

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 24, 2002, 12:10:37 ; Search time 72.239 Seconds  
(without alignments)  
3854.741 Million cell updates/sec

Title: US-09-720-086-5

Perfect score: 4909

Sequence: 1 MFSSGPGDPTSSSSLEREDDR.....MSVPYRHLFAPLKEYFACV 908

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 6.0, Ygapext 7.0  
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n model -DEV=xlh  
-Q=/cgnt2\_1/USFTO.spool/US09720086/runat.18112002.092957.29457/app\_query.fasta.1.4252  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=tni -MINMATCH=0.1 -LOOPL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09720086 @CGN 1.1.108 @runat.18112002.092957.29457 -NCP1=6 -ICPU=3  
-NO\_XLPHY -NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEVTIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Data base:

- 1: Issued Patents NA.\*
- 2: /cgnt2\_6/prodata/1/ina/5A.COMB.seq.\*
- 3: /cgnt2\_6/prodata/1/ina/5B.COMB.seq.\*
- 4: /cgnt2\_6/prodata/1/ina/6A.COMB.seq.\*
- 5: /cgnt2\_6/prodata/1/ina/6B.COMB.seq.\*
- 6: /cgnt2\_6/prodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	748	15.2	2077	4	US-09-276-531-47
2	165	3.4	4190	4	US-08-924-345-1
3	165	3.4	35060	3	US-08-814-095-7
4	165	3.4	43280	4	US-08-804-227C-1
5	162	3.3	4350	4	US-09-295-593-37
6	161	3.3	12001	1	US-08-458-568A-11
7	159.5	3.2	2608	4	US-09-154-750A-75
8	158.5	3.2	11907	4	US-08-061-376-4
9	158.5	3.2	14255	1	US-08-320-559-1
10	158.5	3.2	14255	1	US-08-337-392-1
11	158.5	3.2	14255	1	US-08-306-691B-55
12	158.5	3.2	14255	3	US-08-545-860D-1

13	158.5	3.2	14255	5	PCT-US94-04496-1	Sequence 1, Appli
14	157	3.2	4201	1	US-08-080-255-4	Sequence 4, Appli
15	157	3.2	4201	1	US-08-465-713-4	Sequence 4, Appli
16	157	3.2	4201	5	PCT-US93-05857-4	Sequence 4, Appli
17	153	3.1	11220	4	US-09-105-537-12	Sequence 32, Appli
18	153	3.1	36778	4	US-09-105-537-5	Sequence 5, Appli
19	153	3.1	38506	3	US-09-320-878-19	Sequence 19, Appli
20	152.5	3.1	8438	1	US-07-945-283-1	Sequence 1, Appli
21	151.5	3.1	4852	1	US-07-853-913-3	Sequence 1, Appli
22	151.5	3.1	6755	3	US-08-931-999-4	Sequence 4, Appli
23	150.5	3.1	5267	3	US-08-976-255-2	Sequence 2, Appli
24	150	3.1	2355	4	US-08-913-153-12	Sequence 12, Appli
25	150	3.1	3773	4	US-09-130-242-1	Sequence 1, Appli
26	149	3.0	13842	4	US-09-105-537-30	Sequence 30, Appli
27	149	3.0	36778	4	US-09-105-537-5	Sequence 5, Appli
28	149	3.0	38506	3	US-09-320-878-19	Sequence 19, Appli
29	149	3.0	43280	2	US-08-804-227C-1	Sequence 1, Appli
30	148.5	3.0	1566	2	US-08-405-175A-10	Sequence 10, Appli
31	148	3.0	9551	1	US-08-056-200-93	Sequence 93, Appli
32	148	3.0	9551	2	US-08-800-644-93	Sequence 93, Appli
33	147.5	3.0	2115	3	US-09-032-365A-12	Sequence 12, Appli
34	147.5	3.0	2116	1	US-08-701-380-1	Sequence 1, Appli
35	147	3.0	2680	4	US-09-063-035-1	Sequence 1, Appli
36	147	3.0	15872	4	US-09-105-537-1	Sequence 1, Appli
37	146	3.0	1926	4	US-09-249-585A-2	Sequence 1, Appli
38	146	3.0	2580	3	US-09-050-863-2	Sequence 2, Appli
39	146	3.0	2580	4	US-09-359-081-2	Sequence 2, Appli
40	146	3.0	5452	2	US-09-130-114-1	Sequence 1, Appli
41	146	3.0	9600	4	US-08-910-647-1	Sequence 1, Appli
42	146	3.0	9600	4	US-09-620-925-1	Sequence 1, Appli
43	146	3.0	10596	1	US-07-884-811-15	Sequence 15, Appli
44	146	3.0	10596	1	US-07-885-971-15	Sequence 15, Appli
45	146	3.0	10596	1	US-08-087-783A-15	Sequence 15, Appli

#### ALIGNMENTS

RESULT 1  
US-09-276-531-47  
Sequence 47, Application US/09726531  
Patent No. 6183968  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Xue, Henry  
APPLICANT: Reedy, Koopa  
APPLICANT: Guegler, Karl J.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276,531  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 60/079,677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

```

; NAME: Lynn E. Murty, Ph.D.
; REGISTRATION NUMBER: 42,918
; REFERENCE/DOCKET NUMBER: PA-0008 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2077 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TESTTUT02
; CLONE: 1271435
; US-09-276-531-47

Alignment Scores:
Pred. No.: 3,996-52 Length: 2077
Score: 748.00 Matches: 151
Percent Similarity: 81.95% Conservative: 17
Best Local Similarity: 73.66% Mismatches: 36
Query Match: 15.24% Indels: 4
DB: 4 Gaps: 1

US-09-720-086-5 (1-908) x US-09-276-531-47 (1-2077)
QY 703 GlySerProCysAsnAspLeuSerIleValAlaProAlaArgIleValLeuTyrcIuGly 722
DB 3 GCGCGCANTGCAACAGCTNNAAAT-GTGAATCA-GCCAGGAAGGCTGTATGAGGAC 60
QY 723 ThrGlyArgLeuPheGluPheTyArgLeuLeuHisAspAlaArgProIysGluGly 742
DB 61 AGG---CGGCTCTCTTCGAATTTTC-CACCTGCTGAATCTACCGCCCAAGAGGGT 116
QY 743 AspAspArgProPheTrpLeuPheGluAsnValValAlaMetGlyValSerAspIys 762
DB 117 GATGACCGGCGCTCTTCGAGTGTAAATCCAGTGATGATTCATCAAAAGTTTCT 176
QY 763 ArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaIysGluValSer 782
DB 177 GGGACATCTTCACGGTTCCTGGAGTGTAAATCCAGTGATGATTCATCAAAAGTTTCT 236
QY 783 AlaAlaHisArgAlaArgTy-PheTrpGlyAsnLeuGluGlyCysLeuGluHisGlyValGlyAlaIys 802
DB 237 GCTGCTCACAGGCGCGGATCTTCTGGGGCAACCTACCCGGGATGAACAGGCCCGTGATA 296
QY 803 SerThrValAsnAspIysLeuGluGlnGluCysLeuGluHisGlyValGlyAlaIys 822
DB 297 GCATCAAGATGATAAATCTCGNGCTGCAGGACTGCTTGGATACAAATAGGATAGCCCAAG 356
QY 823 PheSerIleValArgThrIleThrThrArgSerAsnSerIleIysGlnGlyIysAspGln 842
DB 357 TTAAGAAAGTACACAAATACCAACCAAGTCGAACTCGATCAACAGGGGAAACCAAA 416
QY 843 HisPheProValPheMetAsnGluIysGluAspIleLeuTrpCysThrGluMetGluArg 862
DB 417 CTTTTCCTCTGTTGTATGAATGGCAAGAAAGATGTTTNGTGGTGCACTGAGCTCGAAAGG 476
QY 863 ValPheGlyPheProValHisTyThrAspValSerAsnMetSerArgLeuAlaArgGln 882
DB 477 NTCTTTGGTTCCTGTCACATACACAGAGCTGTCCAACATGGCGCGTGGTCCCGCCAG 536
QY 883 ArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAlaProLeuIys 902
DB 537 AAGCTGCTGGGAAGGTCTCTGGAGGCTGCTGTATCCGACACCTCTTCGCCCTCTGAAG 596
QY 903 GluTyPheAlaCys 907
DB 597 GACTACTTTTGCAATGT 611

RESULT 2
US-09-924-345-1

```

```

; Sequence 1, Application US/08924345
; Patent No. 6224878
; GENERAL INFORMATION:
; APPLICANT: LEUNG-TACK Patricia
; APPLICANT: LEGASTELOIS Isabelle, Christine, Marie-Andree
; APPLICANT: AUDONNET Jean-Christophe, Francis
; APPLICANT: RIVIERE Michel, Emile, Albert
; TITLE OF INVENTION: Mutants and vaccines of the Infectious
; TITLE OF INVENTION: Bovine Rhinotracheitis virus
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LARSON AND TAYLOR
; STREET: 727 SOUTH TWENTY-THIRD STREET
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/924,345
; FILING DATE: 04-SEP-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/199,172
; FILING DATE: 09-AUG-1994
; APPLICATION NUMBER: FR 92 07930
; FILING DATE: 26-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: SARRO, THOMAS P.
; REGISTRATION NUMBER: 19396
; REFERENCE/DOCKET NUMBER: XI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 920-7200
; TELEFAX: (703) 892-8428
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4190 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Type 1 Bovine Herpesvirus
; STRAIN: ST
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 172..1311
; OTHER INFORMATION: /function= "envelope glycoprotein"
; OTHER INFORMATION: /product= "glycoprotein g1"
; OTHER INFORMATION: /standard_name= "BHV-1 g1"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1594..3318
; OTHER INFORMATION: /function= "envelop glycoprotein"
; OTHER INFORMATION: /product= "glycoprotein gE"
; OTHER INFORMATION: /standard_name= "BHV-1 gE"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3388..3861
; OTHER INFORMATION: /product= "protein US9"
; OTHER INFORMATION: /standard_name= "BHV US9"
; US-08-924-345-1

Alignment Scores:
Pred. No.: 0,000935 Length: 4190
Score: 165.00 Matches: 119
Percent Similarity: 32.78% Conservative: 58

```



```
,
,
, TELEPHONE: (248) 539-5050
, TELEFAX: (248) 539-5055
, INFORMATION FOR SEQ ID NO: 7:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 35060 base pairs
, TYPE: nucleic acid
, STRANDEDNESS: double
, TOPOLOGY: linear
, MOLECULE TYPE: other nucleic acid
, DESCRIPTION: /desc = "Cosmid including ACH2
, DESCRIPTION: /desc = "Cosmid including ACH2
, HYPOTHETICAL: NO
, ANTI-SENSE: NO
, ORIGINAL SOURCE:
, ORGANISM: Homo sapiens
, POSITION IN GENOME:
, CHROMOSOME/SEGMENT: 7q22
, FEATURE:
, NAME/KEY: promoter
, LOCATION: 4089..22464
, OTHER INFORMATION: /function= "ACHE Promotor"
, OTHER INFORMATION: /standard_name= "ACHE Promotor"
, FEATURE:
, NAME/KEY: exon
, LOCATION: 22465..22537
, OTHER INFORMATION: /function= "non-translated"
, OTHER INFORMATION: /gene= "ACHE"
, OTHER INFORMATION: /number= 1
, FEATURE:
, NAME/KEY: exon
, LOCATION: 24090..25177
, IDENTIFICATION METHOD: experimental
, OTHER INFORMATION: /function= "(translation start:
, OTHER INFORMATION: 24110)"
, OTHER INFORMATION: /evidence= EXPERIMENTAL
, OTHER INFORMATION: /gene= "ACHE"
, OTHER INFORMATION: /number= 2
, FEATURE:
, NAME/KEY: exon
, LOCATION: 25524..26009
, IDENTIFICATION METHOD: experimental
, OTHER INFORMATION: /evidence= EXPERIMENTAL
, OTHER INFORMATION: /gene= "ACHE"
, OTHER INFORMATION: /number= 3
, FEATURE:
, NAME/KEY: exon
, LOCATION: 27005..27274
, IDENTIFICATION METHOD: experimental
, OTHER INFORMATION: /evidence= EXPERIMENTAL
, OTHER INFORMATION: /gene= "ACHE"
, OTHER INFORMATION: /number= 4
, FEATURE:
, NAME/KEY: exon
, LOCATION: 27255..28007
, IDENTIFICATION METHOD: experimental
, OTHER INFORMATION: /evidence= EXPERIMENTAL
, OTHER INFORMATION: /gene= "ACHE"
, OTHER INFORMATION: /number= 5
, FEATURE:
, NAME/KEY: terminator
, LOCATION: 27385..27387
, FEATURE:
, NAME/KEY: exon
, LOCATION: 28008..28129
, IDENTIFICATION METHOD: experimental
, OTHER INFORMATION: /evidence= EXPERIMENTAL
, OTHER INFORMATION: /gene= "ACHE"
, OTHER INFORMATION: /number= 6
, FEATURE:
, NAME/KEY: terminator
, LOCATION: 28129..28131
, FEATURE:
, NAME/KEY: exon
, LOCATION: complement (34528..34895)
, OTHER INFORMATION: /function= "arsenite resistance
, OTHER INFORMATION: /gene=
, OTHER INFORMATION: /gene= "AR"
, OTHER INFORMATION: /number= 1
, FEATURE:
, NAME/KEY: exon
, LOCATION: complement (34092..34358)
, OTHER INFORMATION: /gene= "AR"
, OTHER INFORMATION: /number= 2
, FEATURE:
, NAME/KEY: exon
, LOCATION: complement (33779..33963)
, OTHER INFORMATION: /gene= "AR"
, OTHER INFORMATION: /number= 3
, FEATURE:
, NAME/KEY: exon
, LOCATION: complement (33493..33591)
, OTHER INFORMATION: /gene= "AR"
, OTHER INFORMATION: /number= 4
, FEATURE:
, NAME/KEY: exon
, LOCATION: complement (33297..33408)
, OTHER INFORMATION: /gene= "AR"
, OTHER INFORMATION: /number= 5
, FEATURE:
, NAME/KEY: exon
, LOCATION: complement (32959..33094)
, OTHER INFORMATION: /gene= "AR"
, OTHER INFORMATION: /number= 6
, FEATURE:
, NAME/KEY: exon
, LOCATION: complement (32569..32628)
, OTHER INFORMATION: /gene= "AR"
, OTHER INFORMATION: /number= 7
, FEATURE:
, NAME/KEY: exon
, LOCATION: complement (32386..32468)
, OTHER INFORMATION: /gene= "AR"
, OTHER INFORMATION: /number= 8
, FEATURE:
, NAME/KEY: exon
, LOCATION: complement (31894..32080)
, OTHER INFORMATION: /gene= "AR"
, OTHER INFORMATION: /number= 9
, FEATURE:
, NAME/KEY: exon
, LOCATION: complement (31363..31534)
, OTHER INFORMATION: /gene= "AR"
, OTHER INFORMATION: /number= 10
, FEATURE:
, NAME/KEY: exon
, LOCATION: complement (31131..31284)
, OTHER INFORMATION: /gene= "AR"
, OTHER INFORMATION: /number= 11
, FEATURE:
, NAME/KEY: exon
, LOCATION: complement (30816..31011)
, OTHER INFORMATION: /gene= "AR"
, OTHER INFORMATION: /number= 12
, FEATURE:
, NAME/KEY: exon
, LOCATION: complement (30470..30626)
, OTHER INFORMATION: /gene= "AR"
, OTHER INFORMATION: /number= 13
, FEATURE:
, NAME/KEY: exon
, LOCATION: complement (30187..30274)
, OTHER INFORMATION: /gene= "AR"
, OTHER INFORMATION: /number= 14
, FEATURE:
, NAME/KEY: exon
, LOCATION: complement (29945..30073)
```







Db 22932 CGAGGCGCGGTCAAGTGTCTGTGCTCCGCAACGCTTCAACTCCT---TGAAG 22876  
 QY 282 AAGGAYAGGlypHeGlylleglylueValTroplylueuArgglypHeSerTrp 301  
 Db 22875 GAAGGCGGTGGGGGTTCTGGGCATCGG---TGGGGC-----TTCGGG 22834  
 QY 302 TTPProGlyArggileValSerTrpTrpMetThrglyArgSerArgAlaAGlyl--- 320  
 Db 22833 TGGTCGTCCGGG---GGGTGGTGG---GGGTGTCTCCAGGCACTAGTGTGTGT 22786  
 QY 321 -----ThragTTPValMetTrp-----pHeGlyAspGlyLysPheSerVal 334  
 Db 22785 GGTGGAAGGGGTAGTGGAGGGGTGTGAGTGGCTGGGGGAGGGGGG---TCCAG 22727  
 QY 335 ValCyValGlyLysLeuMetPProLeuSerSerPheCySerAlaPheHISGlnAlaThr 354  
 Db 22726 GTGAGGGTGT----- 22718  
 QY 355 TyzAnLysGlnPProMetTyArgLysAlaIleTyrglyLueGln-ValAlaSerSe 374  
 Db 22717 -----GGCGGTGGGCGCCAGGCTCGGGGTGGAGTCCAGCAGCGGTGGGGCTG 22667  
 QY 374 rArgAlaGlyLysLeuPheProAlaCyHISAspSerArgLysSerArgLysAl 394  
 Db 22666 TCGTCTGTGGCGCGCGAGGTCCGGTGGCAT-----GGACCTCCGGTCCGGC 22619  
 QY 394 aValGlyValGlnAsnLysGlnMetIleGlyTTPAlaLueGlyGlyPheGlnProSeG 414  
 Db 22618 GCCTCGG---CCGTCTGTCTCAATGGCATGMAAGAGTACCGGTGGGCACTGACCTCGATG 22562  
 QY 414 YProLysGlyLueGlnProProGlyLueGlyLysAsnProTyrglyLueValTyThrAs 434  
 Db 22561 AACGAGGTGGCGCTCTCCCAACAACACCGACCCGCGAGAGA----- 22516  
 QY 434 pMetTyrglyLueGlnProGlyLueAlaAlaTyrglyLueProProProProAlaLysPProAr 454  
 Db 22515 -----ACTCAACCGGTTCGCGAGATTACGMAACAATACCCGCGATCGAAGACCGC 22463  
 QY 454 GlyS 455  
 Db 22462 TCGC 22459  
 RESULT 5  
 09-295-593-37/c  
 Sequence 37, Application US/09295593  
 Patent No. 6417169  
 GENERAL INFORMATION:  
 APPLICANT: WRIGHT, Jim A.  
 APPLICANT: YOUNG, Aiping H.  
 APPLICANT: LEE, Yoon S.  
 TITLE OF INVENTION: INSULIN-LIKE GROWTH FACTOR II ANTISENSE OLIGONUCLEOTIDE  
 TITLE OF INVENTION: SEQUENCES AND METHODS OF USING SAME TO MODULATE CELL  
 FILE REFERENCE: 032396-046  
 CURRENT APPLICATION NUMBER: US/09/295,593  
 EARLIER FILING DATE: 1999-04-22  
 EARLIER APPLICATION NUMBER: US 60/082,791  
 NUMBER OF SEQ ID NOS: 37  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 37  
 LENGTH: 4350  
 TYPE: DNA  
 ORGANISM: Human  
 US-09-295-593-37  
 Alignment Scores:  
 Pred. No.: 0.00174  
 Score: 162.00  
 Percent Similarity: 32.93%  
 Best Local Similarity: 21.00%  
 Query Match: 3.30%  
 DB: 4  
 Gaps: 30

US-09-720-086-5 (1-908) x US-09-295-593-37 (1-4350)  
 QY 16 ArgGlyAspAspArgglypGlyLueGlyLueGlnGlyLueGlnAsnArgglyLys-----Gln 33  
 Db 3519 AGAGCTTCGAGCTCTCTGAGCAGACGAGGTCTGAGGGGGGGGCAAGATCTCACCGA 3460  
 QY 34 GluArgGlnGlyProSerAlaThrAlaArgLysVal----- 45  
 Db 3459 GGGAGAGGGAGGGTCTCTCAAGTGTGCGAAGCGGCCACCTGGCCAGTATGAA 3400  
 QY 46 -----GlyArgProGlyArgLys 53  
 Db 3399 CCCTGTGTTACACACTGTAGCCCTTCCCTCCGCGCAGCTAGCCAGCAGAG 3340  
 QY 54 HisProValGlyLysSer-----AspThrProLysAspProAlaValThrThr 70  
 Db 3339 CCCCCTCTGTGGCTTCGAGCACCTCTGACACCTGAGAGACCTTGCCTCA-ACCC 3281  
 QY 71 LysSerGlnProMetAlaGlnAspSerGlyProSerAspLeuProAsnGlyAspLeu 90  
 Db 3280 -----CCCTCCGCTCAGACCTTGGGGACGAG 3251  
 QY 91 GlyLysArgSerGlyProGlnProGlyLueGlySerPro-----AlaAla 105  
 Db 3250 CCAGTGTCCCGAAGCGAGCGGACATGCTTCTCCACTCCCTGTGACTTCTCAAG 3191  
 QY 106 GlyLysGlyGlyAlaProAlaGlyLueGlyLueGlyLueGlyLueGlyLueGlyLue 125  
 Db 3190 GGGGCTCATGCGCTGCCCC-----CAGAGGCTCCACCTCAACT--- 3149  
 QY 126 ArgAlaValGlnAsnGlyCyCyValThrLysGlyLueGlyAlaSerAlaGlyLue 145  
 Db 3148 -----CAA 3146  
 QY 146 GlyLysGlyLueGlyLueGlyLueGlyLueGlyLueGlyLueGlyLueGlyLueGlyLue 165  
 Db 3145 CCGAAGCAAGCGAGCAG-----ATGAACCAACCAATCCAAATCAGGCGCA 3101  
 QY 166 ---LeuArgGlyLueGlyLueGlyLueGlyLueGlyLueGlyLueGlyLueGlyLue 184  
 Db 3100 TAAATGCGGGGCGAGACGAGTGTCTCCAGGCTGCTTCTGCTTCTTGG 3041  
 QY 185 Thr-----PheGlnAlaGlyAspProTyrglyLysSerLysArgLysArgAsp 200  
 Db 3040 TCACTGCCCCCTGTACATGGGGGGGGTTAATTGCTTCTGAGCGCATAAAG--- 2984  
 QY 201 GlyTTPLeuAlaArgTyrglyLueGlyLueGlyLueGlyLueGlyLueGlyLueGlyLue 215  
 Db 2983 -----CTAAGAGGGGTAAAAAACAACAAAAAATAAAGAAAAATGCCCCA 2931  
 QY 216 -----ValIleAlaVal 219  
 Db 2930 AGAACCAAGAGGGGATATGATCATGATGAATTTGGGGTTCTTAATTTCTTA 2871  
 QY 220 MetAsnAlaVal-----GlnGlu 225  
 Db 2870 TGTAAATTCGCAAGTGTCTGATCATCTCTGACTTTCCATCCAAATCTCCGGGAC 2811  
 QY 226 AsnGlnAlaSerGlyLueSerGlyLueGlyLueGlyLueGlyLueGlyLueGlyLueGlyLue 244  
 Db 2810 CACTTCTTACCCAGAACTCCCGCGCCAGCCCGACCTGCCCCCAGCCACCTCCCA 2751  
 QY 245 GlnProThrAspProAlaSerProThrValAlaThrThrProGlyPro-----Val 264  
 Db 2750 CCGCCAGACTTCCCACTCCCGCATCACTGAGCGCCCGCCAGCTCCCACTTCTT 2691  
 QY 262 GlyGlyAspAlaGlyAspLysAsnAlaThrLysAlaAlaAspAspGlyProGlyTyrgly 281  
 Db 2690 GGGGGTCCCGAGGA----- 2676  
 QY 282 AspGlyArgGlyPheGlylleglylueValTroplylueuArgglypHeSerTrp 301



```

QY 95 -----GluProGlnProGlnGluGly 101
Db 6162 GGTGGGGTGTGTGCGGGGTGTCCCGCCTCTCCCTCTCCGCGGCCCGCCCGACCGAGGG 6221
QY 102 SerPro-----AlaAlaGlyGlnLysGlyGlyAlaPro 112
Db 6222 CCCCCCTCTGTCGGCTGTGGCTGGGGTGGGGCGCGCGCGCGCGCGCGCGCGCGGG 6281
QY 113 AlaGlyGlyGlnGlyThrGlnThrProProGlnAlaSerArgAlaValGlnGlnGlyCys 132
Db 6282 AGCGGGGGCGCTGTTCTTCTCCGACCGCATCGCCGATGCGGGCGCATCTCCGGGAGATA 6341
QY 133 CysVal-----ThrysgGlnGlyArgGlyAlaSerAlaGlyGlu---GlyLysGlu 148
Db 6342 CGCTCGACGCGCGCGACGTACGAGGTAGTCACTTACGACTCTGATGGGGGAGGG 6401
QY 149 GlnLysGlnThrAsnIleGlnSerMetLysMetGlnGlySerArgGlyArgLysAlaArgGly 168
Db 6402 GCGAGACCCACGAGACCCCGACGACCCCGCTCGACCGGAACTAGCGCGACCGGTG 6461
QY 169 GlyLeuGlyTyrPheLysSerLeu---ArgGlnArgProMetProArgLeuThrPheGln 187
Db 6462 ATGCTTGGGGGGGAAAAAGACAGAGAGCGCGCATCCCTCCGCGCTTCTGTC----- 6515
QY 188 AlaGlyAspProTyrTyrIleSerLysArgLysArgAspGlu----- 201
Db 6516 -----CGCGATCGCGCTCCCGCGCGCGAGCGCTCTGACGCTGTCTCTC 6560
QY 202 TrpLeuAlaArg-----Trp-----LysArgGlu 209
Db 6561 TGGCGGTCCGCGCTGGGTCTGTGATCCGTCGCGACCGCGCTCCGTGGAGCGATCG 6620
QY 210 AlaGlnLysValAlaValIleAlaValMetAsnAlaValGlnGlnAlaSer 229
Db 6621 GCGGCTCTCGGCTCATATATGTC-----CCA 6647
QY 230 GlyLysSerGlnLysValGlnGlnAlaSerProAlaValGlnGlnProThrAspPro 249
Db 6648 GGGCGCGCGGAGAGAGAGACGCGAGCGCGCGCG-----CCCCGCCCCC 6698
QY 250 AlaSerProThr-----ValAlaThrThr----- 257
Db 6699 GCGCGCGCCACCCCGACGAAATTCATTATGACGACCGCCCGCCGCGCGACGCC 6758
QY 258 -----ProGlnProValGlyLysAspAlaGlyAspLysAsnAlaThrLysAla 273
Db 6759 GGGGGCGCGTGGCGCGCGCGCGCTGTGTAACCGCGCGCGCGCGCAT----- 6806
QY 274 AlaAspAspGlnProGlnTyrGlnAspGlyArgGlyPheGlyIleGlyGlnLeuValTrp 293
Db 6807 -----CCGCGCATCTGCCA----- 6821
QY 294 GlyLysLeuArgGlyPheSerTrpTrpProGlnArgGlyLeuValSerTrpTrpMetThrGly 313
Db 6822 -----TGGGCGGGGCGC-----GAGCGCGGTGGGTCCGCGCC 6854
QY 314 ArgSerArgAlaAlaGlnGlyThrArgTrpValMetTrpPheGlyAspGlyLysPheSer 333
Db 6855 CCGCCCCCGATGCGCATCTCATACGC----- 6881
QY 334 ValValCysValGlnLysLeuMetProLeuSerSerPheCysSerAlaPheIleGlnAla 353
Db 6882 -----CCGATCCGCGGCTTCCGCTTCCGCTTCCGCGCATGCTA 6917
QY 354 ThrTyrAsnLysGlnProMetTyrArgLysAlaIleTyrGlnValLeuGlnValAlaSer 373
Db 6918 ACGAGAGACGCGCGAG-----GGG 6935
QY 374 SerArgAlaGlyLysLeuPheProAlaCysHisAspSerAspGlnSerAspSerGlyLys 393
Db 6936 GCGCGCGCGCGCGCGCGCATCTCC-----GGTTCG 6965

```

```

QY 394 AlaValGlnValGlnAsnLysGlnMetIleGlnTrpAlaLeuGlyPheGlnProSer 413
Db 6966 GCGGTAAATG-----AGATACGAGCCCGC 6989
QY 414 GlyProLysGlyLeuGlnProProGlnGlnGlyLysAsnProTyrLysGlnValTyrThr 433
Db 6990 GCGCGCGTGGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7031
QY 434 AspMetTrpValGlnProGlnAla-AlaAlaTyrAlaProPro-----ProProAlaLys 451
Db 7032 GACCGCGGAGACCAAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 7091
QY 451 LysProArgLysSerThrThrGlnLysProLysValLysGlnIleLeuAspGlnArgTh 471
Db 7092 TTGGCGCGCGCGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7151
QY 471 Arg-----GlnArgLeuValTyrGlnValArgGlnLysCysArgAsn 485
Db 7152 GCGAAGCGTTGCACTTCTCCCATATATATATATTATTAGGGCGAAGTGGAGAC 7209

```

## RESULT 7

```

US-09-154-750A-75
; Sequence 75, Application US/09154750A
; Patent No. 6412640
; GENERAL INFORMATION:
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Polyak, Kornelia
; TITLE OF INVENTION: p53-Induced Apoptosis
; FILE REFERENCE: 1107/75357
; CURRENT APPLICATION NUMBER: US/09/154,750A
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/059,153
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/079817
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 75
; LENGTH: 2608
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-154-750A-75

```

## Alignment Scores:

```

Pred. No.: 0.00132 Length: 2608
Score: 159.50 Matches: 153
Percent Similarity: 32.04% Conservative: 79
Best Local Similarity: 21.13% Mismatches: 278
Query Match: 3.25% Indels: 217
DB: Gaps: 33

```

US-09-720-086-5 (1-908) x US-09-154-750A-75 (1-2608)

```

QY 4 SerGlyProGlyAspThrSerSerSerLeuGlnArgLysAspAspArgLysGlnGly 23
Db 632 TCTGGCCAGC-----TGGCGAGCATTCACGACGACCTGAAGGG 670
QY 24 GlnGlu-----GlnGlnGlnAsnArgGlyLysGlnGlnArgGlnGlnPro 38
Db 671 GAGGATCTCGGACTCGAGTGGAGCGGTAGTACAGACAGAGTGGCGGTGAAGCCCTGGA 730
QY 39 SerAlaThrAlaArgLysValGlyArgProGlyArgLysArgLysHisProProValGln 58
Db 731 GTGGAGAGAGAGAGAGCGTGCAGAGACAGCTGACCTGCAACAAAAAATGCCAGAAC 790
QY 59 SerSerAspThrProLysAspProAla---ValThrThrLysSerGlnProMetAlaGln 77
Db 791 CAATCAGAAATTCAGAAATCCAGAGCCAGGCAAGAGACAGACAGACAGACAGACCA 850
QY 78 Asp-----SerGlyProSerAspLeuLeuProAsnGlyAspLeuGlnLysArgSerGlu 95
Db 851 GATTAAAGATCTGAGAGAGAACTGTCTCTGCAAGAGCAGAGATGACGATTTGAAGAA 910

```

QY	96	ProGlnProGluGluGly---SerProAlaAlaGlyGlnLysGlyGlyAlaProAlaGlu	114
Db	911	CATGAGTCTGAGCTGGTACGCTCCCTAGCTGGAAACGGGAGCTGGAGCAGCTGGCGGA	970
QY	115	GlyGluGlyThrGluThrProGluAlaSerArgAlaValGluAsnGlyCysCysVal	134
Db	971	GGAGAGCGCACTCGGGAGATCAGAGAGACCAACGGGCTGCTCCAGGA	1018
QY	135	ThrLysGluGlyArgGlyAlaSerAlaGlyGluGly-LysGluGlnLysGlnThrAsnIl	154
Db	1019	---AGAGCTGGAAGGCTCAGAGAGAACTGGGCGCCAGAGAGAATCGAGAGAGCGCT	1075
QY	154	eGluSerMetLysMetGluGlySerArg-----GlyArgLeu-ArgGlyGlyLeuGlyT	172
Db	1076	GGTTGGCTTGGAGCTGGAGACGAGAGAGCTGCTGGCCAAGCTGCAAAAGCTGGAGAGACT	1135
QY	172	rpGlu-----SerSerLeuArgGlnArgProMetProArgLeuThrPheG	187
Db	1136	GGACCAGACCATGGGCGCTTGAGCATCAGAGACTCCAGAGAAGACCTTCCA	1182
QY	187	InlaGlyAspProTyrTyrIleSerLysArgLysArgAspGluTrpLeuAlaArgTrpL	207
Db	1183	-----GATTCTGGTGGTTCAGCTG-CAGC	1203
QY	207	ysArgGluAlaGluLysLysAlaLysValIleAlaValMetAsnAlaValGluGluAsnG	227
Db	1204	AGAGGAGCTGCTTGAAGACAGCAAGACAGCGCGCTCAC-----CAGCAGCGCC	1253
QY	227	InlaSerGlyGluSerGlnLysValGluGluLaseProProAlaValGlnGlnProT	247
Db	1254	CGGGGCTGGAGAAGCCAGGCAGCAGCTGCAGGAGGAGCTCCGGCAGGTCACGCGCCAG	1313
QY	247	hrAspProAlaSerProThrValAlaThrPrGluProValGlyGlyAspAlaGlyA	267
Db	1314	CT-----GTTGGAGGAGAGAGAGAAG	1334
QY	267	spLysAsnAlaThrLysAlaAlaAspAspGluProGluTyrGluAspGlyArgGlyPheG	287
Db	1335	CGCAGACCCACGAGGCGCTGGCCGCGAGGCTCCAGAAACGGGTCTGTCTGTCTACCAAG	1394
QY	287	lyIleGlyGluLeuValTrpGlyLysLeuArgGlyPheSerTrpTrpPro-----	303
Db	1395	GAGCG-----GGACGTATGCGGGCCA-TCCTGGGGTCTCTACGACAGCGA	1438
QY	304	-----GlyArgIleValSerT	309
Db	1439	GCTGACCCCGGCGGCTACTCTACCCACGCTACCGCGCGCATCGGGAGGCTGAGGAT	1498
QY	309	rp-----TrpM	311
Db	1499	GGTCAGAAAGTCCACAGCCAGCCGAGATGGAGGCTCAGCTGTCCGAGGCCCTCGA	1558
QY	311	etThrGlyArgSerArgAlaGluGlyThrArgTrpValMetTrpPheGlyAspGlyL	331
Db	1559	GGAGCTGGAGGCGCCAGAAACAAAGAGCAGATCTGCTG-----A	1597
QY	331	ysPheSerValValCys-----ValGluLysLeuMetProLysSerSerPheCysSerA	349
Db	1598	GATGGAGCTGGAAGATGCTGAAAGTCTCAGTCCAGCTCTGCCAACAGAGCTCTCTGTTCTC	1657
QY	349	laPheHisGlnAlaThrTyrAsn-----	356
Db	1658	CAGGGAGGAGCGGACACGCTCAGTTGAAGGTCGAGGAGCTGGNAGCGAGCGAGTCG	1717
QY	357	-----LysGlnProMetTyrArgLysAla---IleTyrGluValLeuGlnValA	372
Db	1718	GCTGGAGGAGGAAAAAGAGGATGCTGGAGGCACAGCTGGAGCGGCGAGCTCTGCGAGGTGA	1777
QY	372	laSerSerArgAlaGlyLysLeuPheProAlaCys-HisAspSerAspGluSerAspSer	391
Db	1778	CTATCACCAGACGAGGCCA-----AAGTGTGCACATGAGCTGAAACCCACCAAGTCT	1831

Qy	392	GlyLysAlaValGluValGlnAsnLysGlnMetIleGluTrpAlaLeuGlyGlyPheGln	411
Db	1832	GGCCAGGACGCGCTCGCGAGGACCAAGCCA	1882
Qy	412	ProSerGlyProLysGlyLeuGluProProGluGluGluLysAsnProTyrLysGluVal	431
Db	1883	GCAGACTCGCGGGCTCTCGCGCCATGGAGAGGAGGACCGTCCACGCCGA	1936
Qy	432	TyrThrAspMetTrpValGluPro-GluAlaAlaTyrAlaProProProPro	449
Db	1937	---CCTTGAGGCTCGCGCGGAGTCTGCCATCGTCCAAGGA	1975
Qy	450	---AlaLysLysProArgLysSerThrThrGluLysProLysValLysG1	465
Db	1976	GGTGGCAGAGCTGAAGAGACAGGTGGAGAGTGGCGAGCTGAAGAAGAACCAACGCGCTCAAGGA	2035
Qy	465	u1lelleAspGluArgThrArgGluArgLeuValTyrGluValArgGlnLysCysArgAs	485
Db	2036	GGTTTTCAGACCAAG---ATCCAGGAGTTCCGCAAGCGCTGTACAC	2080
Qy	485	n1leGluAspIleCysIleSerCysGlySerLeuAsn---Valth	499
Db	2081	GCTCACCGGCTACCAGATCGACATCACACGAGGAACAGTACCGGCTGACCTCGCTGTGA	2140
Qy	499	r1leuGluHisProLeuPheIleGlyGlyMetCysGlnAsnCysLys---AsnC	516
Db	2141	CGCCGAGCACCACA---GGCGACTGCTCATCTTCAAGGCCACCAAGCCCTCG	2188
Qy	516	sPheLeuGluCysAlaTyr---GlnTyrAspAspAspGlyTyrGlnSerTyrCysTh	534
Db	2189	GGTTTCCAGATGTCAGCTACTCGGAGACAGAGTTCTCACACCGTGGCGGAGCTCATCGAG	2248
Qy	534	r1leCysCysGlyGlyArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPh	554
Db	2249	GTGCACCTCGCGGCCAGGACAGCATCCCTGCTCTCTCAGCTCGC---2294	
Qy	554	eCysValGluCysValAspLeuValGlyProGlyAlaAlaGlnAlaIleLysG1	574
Db	2295	---TCACCTCGAGCTCTTCAGCGCC---2318	
Qy	574	uAspProTrpAsnCysTyrMetCys---GlyHisLysGlyThrTyrGlyLeuLeuArgAr	593
Db	2319	-AGACCGTGGCTGACCTGCGAGCTCGGGGGCATAGC---CGGAG	2359
Qy	593	gArgGluAspTrpProSerArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPh	613
Db	2360	CCACTCTGCTGGCTCTGAC---CTGCGAGTCT---2387	
Qy	613	eAspProProLysValTyrProProValProAlaGluLysArgLysProIleArgValLe	633
Db	2388	---CCCTGCCCGCCAGCCACAGCTGGGTGGCAGCTCTCTGC	2425
Qy	633	uSer	634
Db	2426	CTCT	2429

RESULT 8  
US-08-061-376-4  
; Sequence 4, Application US/08061376  
; Patent No. 6175000  
; GENERAL INFORMATION:  
; APPLICANT: Evans, Glen A.  
; APPLICANT: Djabali, Malek  
; APPLICANT: Selleri, Lucia  
; APPLICANT: Parry, Pauline  
; TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11Q23  
; TITLE OF INVENTION: TRANSLOCATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pratty, Schroeder, Brueggemann & Clark  
; STREET: 444 South Flower Street, Suite 2000  
; CITY: Los Angeles  
; STATE: California



```
Db 4008 AGGTCAGAGCAGAGCAAGAAAGAGTGGCTCCCGCCCAAGTATCCCTGTTAAACA 4067
Qy 459 ULVSPROLYVALVLSGLU-
Db 4068 AAAACCAAAAGAAAGAAACCAACCTCCGGTCAATAAGCAGGAGAATGCAGGCATTT 4127
Qy 466 ----IleIleAspGluArgThrArg-
Db 4128 GAACATCCTCAGCAGCTCTCTCAATAGGCAATAGTTCTAAGCAAAAAATTCACAGAGATGG 4187
Qy 473 ----GluArgLeuValThrGluValArgGlnLysCys-
Db 4188 AGTCCACAGCATCAGAGTGGAGCTTTAAGGAGGATTTGAAGCAGAGAAATGTTGGGAGAT 4247
Qy 489 ----
Db 4248 GGGAGGCTTAGGAATCTTGACTTCTGTTCTTCCCTATAACACCCAGGGTGGTTGCTTCTCTG 4307
Qy 493 sGlySerLeuAsnValThrLeuGluHisProLeuPheIleGlyMetCysGlnAsnCY 513
Db 4308 TGCCAGTAGTGGG-
Qy 513 s-
Db 4350 TTGTGAGCCCTCCACAAGTTTGTGTTAGAGGAGAACGAGCGCCCTCTGGAGGACCAT 4409
Qy 528 YTYrGlnSerTYrCysThrIleCys-
Db 4410 GGAATAATTGGTGTGTCGTGTTGCAAAATCTGTCACTGTTGTGGAAGCAACATCAGGC 4469
Qy 539 -GlyArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCY 558
Db 4470 TACAAAGCAGCTCTCGAGTGT-
Qy 558 sValAspLeuValGlyProGly-
Db 4524 C-
Qy 577 pAsnCySerMetCysGlyHisGlyThrTYrGlyLeuLeuArgGlyArgGluAspTYr 597
Db 4572 GATCTGTACCAAGTGTGTCGTGTAAAGAGCTGTGGATCCACAACCTCCAGGCAAAAGGTG 4631
Qy 597 pProSerArgLeuGlnMetPhePheAlaAsnAsnHisAsp 610
Db 4632 GGATGCACAGTGGTCTCATGATTTCTCACTGTGTCTATGAT 4671

RESULT 9
US-08-320-559-1
; Sequence 1, Application US/08320559
; Patent No. 5633135
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canani, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
; TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
; TITLE OF INVENTION: All-1 Region
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/320,559
```

```
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/062,443
; FILING DATE: 14 MAY 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/971,094
; FILING DATE: 30-OCT-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,830
; FILING DATE: 27-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/805,093
; FILING DATE: 11-DEC-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-0855
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14255
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ANTI-SENSE: No
; US-08-320-559-1

Alignment Scores:
Pred. No.: 0.0195 Length: 14255
Score: 158.50 Matches: 148
Percent Similarity: 32.17% Conservative: 101
Best Local Similarity: 19.12% Mismatches: 275
Query Match: 3.23% Indels: 254
DB: 1 Gaps: 33

US-09-720-086-5 (1-908) x US-08-320-559-1 (1-14255)
Qy 10 SerSerSerSerLeuGluArgGluArgPheArgGlyGluGluGluGluGluAsn 29
Db 2404 GCTGCAAGAGCGTGGAGAGGCAAGAGTAGAGAGAGAGAGAGAGAGAGAGAG 2463
Qy 30 ArgGlyLysGluGluArgGluGlu-
Db 2464 AATAAGCGGAGTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2523
Qy 41 ThrAlaArgLysValGlyArgProGlyArgLysArg-
Db 2524 GCTTTGTATCTCTGGTGGTGGTTTCCAAAGAGAGAGTGTGGTGAAGATGTTGCCACT 2583
Qy 60 SerAspThrProLysAspProAlaValThrLysSerGlnProMetAlaGlnAspSer 79
Db 2584 TCATCTTCTGCCAAA-
Qy 80 GlyPro-
Db 2638 GGGACTGATATTACTTCTGTGACTCTTTGGGATACAAACAGCTGT-CAAAACCAAAATACT 2696
Qy 93 ArgSerGluProGlnProGluGluGlySerProAlaAlaGlyGlnLysGlyAlaPro 112
Db 2697 TATAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2756
Qy 113 AlaGluGlyGluGlyThrGluThrProPro-
Db 2757 ATCCCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 122
Qy 123 -
Db 2817 TTCCACTTCTCCATAGGCTCCATGTTGGCTCAGGAGAGAGAGAGAGAGAGAG 129
Qy 130 AsnGlyCysCysValThrLysGluGlyArgGlyAla-
Db 141
```



Db	2877	GAGGTTTGCCAGCCTCTCTAA	AAAAAGGCCAAAGCTACGCTCTCCAGATTGAGAAAGATTA	2936
QY	142	-----	-----SerAlaGlyGlnGlyLeu-----GlnGlnLeuSerGlnTh	152
Db	2937	GAGCTTTAAACAAAC	CCAGCACCAGCCCAAGACAGCGGTCAAGAAATGACTCATCAGAGAC	2996
QY	152	AsnAlIeGluSerMetLeuGlnSerAlaGlyArgLeuArgGlyGlyLeuGlyTr		172
Db	2997	CTCTGTGGAGAGAC	CCCGCGATTAACATGCTCTGCAGAAAGACAGCTGTGGCCCTTGGCGCG	3056
QY	172	pGluSerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTy		192
Db	3057	AAAACGAGCTGTGTTTCTT	CTGATGACATGCCACCCTGAGT-----	3096
QY	192	rTyTrIleSerLysArgLysArgAspGluTrpLeuAlaArgTrpLysArgGluAlaGlyLeu		212
	3097	-----	-----GCCTTACCATGG-----	GAAGA 3113
QY	212	sluValAlaLysValIleAlaValMetAsnAlaValGlnGluAsnGlnAlaSerGlyLeu		232
Db	3114	ACGAGAAAAAGTTTTTGCTTCCATGGGGAATGATGCAAGTATCATCATTTGGCTCA-		3171
QY	232	rGlnLysValGlnGluLysSerProPheAlaValGlnGlnProThrAspProAlaSerPr		252
Db	3172	-----GAAATGTGTGAACCTCTTGCT-----	CCACCATCAAAACAATTAAACC	3215
QY	252	oThrValAlaThrThrProGluProValGlyGlyAspAlaGlyAspLysAsnAlaThrLys		272
Db	3216	TGTC-----	-----ACTAGAA	3227
QY	272	sAlaAlaAspAspGluProGluTrpGluAspGlyArgGlyPheGlyTleGlyGlnLeuVal		292
Db	3228	CAGGCAACCCAGAGAACCTCAGATMAAGAAAGACGCG-ATC-----	GAAGCGGT	3277
QY	292	lTrpGlyLysLeuArgGlyPheSerTrpTrpProGlyArgGlyIleValSerTrpTrpMetTh		312
Db	3278	GTTGGCACTGTCCTCGG-----	CTGCCAGTGCTTGAGACCTGCTGTTGTTGTA	3325
QY	312	rGlyArgSerArgAlaAlaGlnGlyThrArgTrpValMetTrpPheGlyAspGlyLysPhe		332
Db	3326	CTAATTGCTAGTAAAGCCAA-----	GTTGGTGTGCCAATTATTAAGA	3370
QY	332	eSerValValCysValGlnLysLeu-----MetProLeuSerSerPheCysSerAlaPhe		350
	3371	AGCACTGCTCGCAAGATGAGAAATGTCAGATCTACATCAATGATGCTTCCAAAGCCTACC		3430
QY	350	eHisGlnAlaThrTrpAsnLysGlnProMetTrpArgLysAlaIleTyrGlnValLeuGln		370
Db	3431	TGCAGAACCMAGCTMAACTGTGAAA-AGAAAGAGAAAAAGTCTAAGACAGTGAAAG		3489
QY	370	nValAlaSerSerArgAla-----		376
Db	3490	AAAGACAGCAAAAGAGACAGTGTGTGAAGACGTGGTGCATCTAGACAAACCTACC		3544
QY	377	-----GlyLysLeuPheProAlaCysHisAspSerAspGluSerAspSerGlyLys		393
Db	3550	CCATCAGCAAGAGAGATCTGTCCCAAG-AAAAAGCAGTAGAGCTCTCCAGAA		3608
QY	393	sAlaValGluValGlnAsnLysGlnMetIleGluTrpAlaLeuGlyGlyPheGlnProSe		413
Db	3609	GCCCGTCGAGAAAGATGAGAA-----	GGGAAATGTCGCGCCC	3650
QY	413	rGlyProLysGlyLeuGlu-----		419
Db	3651	TGGGCTGAATCAAAACAGGCCACCATCTCCAGCTTCCAGAGAAATCAAGACAGAGTCTC		3711
QY	420	-----ProGluGlnGluLysAsn-----ProTyrLysGluVal		431
Db	3711	CCAGCAGCACTGTATCCCGGCTCAACCACTACAGAGACCGCAAGAAAAAGAT		3777
QY	431	lTyrThrAspMetTrpValGluTrpGluAlaAlaTyrAlaProProProProAlaLys		451

```

Db 3771 TCCAAAACACCTCTAGTAGAGCCCAAGAAAAG-----CAGCTTCACCAACCAAGATC 3824
Oy 451 slypPtoArylySer-----ThrThgI 459
Db 3825 AGGTTCACAGACAGAGCAACAGAAAAAGTGCCTCCCGCCCAAGTATCCTGTAAACA 3884
Oy 459 ulysPtoLyseVallysglu-----465
Db 3885 AAAACCAAAAAGAAAAGAAAACCACTCCGGTCATTAAGCAGGAATGACGACCTT 3944
Oy 466 ----lIeIleAspGluArgThrArg-----472
Db 3945 GACATCCTCAGCAGCTCTCTCCAATGAGCAATAGTTCTAAGCAAAAATTCACAGCAGATGG 4004
Oy 473 ----GluArgleuValTyxGluValArgGlnuScys-----ArgAsnIleGluAsp- 488
Db 4005 AGTCCACGAGATCAGACTGAGCTTAAAGAGATGTGAMAGCAAAAAATGTGTGGAAAT 4064
Oy 489 -----lIeCysIleSerCy 493
Db 4065 GGGAGGCTTAGGAATCTTGACTTCTGTCTATACACCCAGGAGGTGTTCCTTCTGTG 4124
Oy 493 sglYserIeAsnAnvalThrIleuGlnHisProleuPheIleGlyIleMetCysGlnAsnCy 513
Db 4125 TGCCAGTAGTGGG-----CATGAGAGTTGTGG-----TATTGCCAGATGCTG 4164
Oy 513 E-----LysAsnCySpheluGluCySalAlyrGlnTyrAspAspAspGI 528
Db 4167 TTGTAGAGCCCTTCACCAAGTTTTGTTTAGAGGAAAGAGCCCTCTGGAGACCACT 4224
Oy 528 yTyxGlnserTyrcysThrIleCys-----CysGly- 538
Db 4227 GGAATATTTGGTTTGTCTGCTGTCGAATTTCTGTCACTTTGTGAAAGCAATCAGAGC 4284
Oy 539 -GlyArgGluValIleuMetCysGlyAsnAsnAsnCySarCySphCySvalGluCy 558
Db 4287 TACAAAGACGCTGTGGAGTGT-----ATAAGTCCCAAGACGTATCACCCTGAGAG 4340
Oy 558 sValAspIleuValGlyProGly--AlaAlaGlnAlaAlaIleYsgIuAspProTr 577
Db 4341 C-----CTGGACCAAACTACCCCAACCAACCAAGAAAGAAAGAAAGTCTG 4384
Oy 577 pasncysTyrcMetCysGlyHislysgIyThrTyrglyLeuIeuAspArgArgGluAspTr 597
Db 4389 GATCTGTACCAAGTGTGTGCTGTGAAGACTGTGATGCCAATCCACGCAAGAGGTG 4444
Oy 597 proSerArgleuGlnMetPhePheAlaAsnHisAsp 610
Db 4449 GGATGCACAGTGTCTCATGATTTCTCAGCTGTGTCAATGAT 4488

RESULT 10
US-08-327-392-1
Sequence 1, Application US/08327392
Patent No. 5633136
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
TITLE OF INVENTION: Canaan1, Eli
TITLE OF INVENTION: ALL-1 Polynucleotides and Monoclonal
TITLE OF INVENTION: Antibodies for Leukemia Detection and
TITLE OF INVENTION: Treatment
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633336r18
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 5.1

```

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/327,392  
FILING DATE: 07/09/92  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 07/971,094  
FILING DATE: 30-OCT-92  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 07/888,830  
FILING DATE: 27-MAY-92  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 07/805,093  
FILING DATE: 11-DEC-91  
ATTORNEY/AGENT INFORMATION:  
NAME: DeLuca Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: TJU-1331  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14255  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
ANTI-SENSE: No  
US-08-327-392-1

Alignment Scores:  
Pred. No.: 0.0195 Length: 14255  
Score: 158.50 Matches: 148  
Percent Similarity: 32.17% Conservative: 101  
Best Local Similarity: 19.12% Mismatches: 275  
Query Match: 3.23% Indels: 254  
DB: 1 Gaps: 33

US-09-720-086-5 (1-908) x US-08-327-392-1 (1-14255)

```
QY 10 SerSerSerLeuGluArgGluAspArgGlyGluGluGluGluGluGlu 29
DB 2404 GCTGACAGAGCGTGGAGAGGACAGAGTAGAGAGAGACCGGGAGAGAGAGAG 2463
QY 30 ArgGlyGluGluArgGluGluGluGluGluGluGluGluGluGluGluGlu 40
DB 2464 AATAAGCGGGAGTCAAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2523
QY 41 ThrAlaArgLysValGlyArgProGlyArgGlyArgGlyArgGlyArgGly 59
DB 2524 GCTTTGTATCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2583
QY 60 SerAspThrProLysAspProAlaValThrLysSerGlnProMetAlaGlnAsp 79
DB 2584 TCATCTTCTGCCAAA-----AAGCAACAGGGCGGAGAGAGTCTTCATCAGATGTTCT 2637
QY 80 GlyPro-----SerAspLeuLeuProAsnGlyAspLeuGluLys 92
DB 2638 GGGACTGATATTACTCTGTGACTCTTGGGATACACAGCTGT-CAAAACCAAAATACT 2696
QY 93 ArgSerGluProGlnProGluGluGlySerProAlaAlaGlyGlnLysGlyAlaPro 112
DB 2697 TATAAAGAAAGGAGAGAAATCTGGAAAAAACCAACTTGGACCTCGGCCCAACTGCCCC 2756
QY 113 AlaGluGlyGluGlyThrGluThrProPro----- 122
DB 2757 ATCCCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2816
QY 123 -----GluAlaSerArgAlaValGlu 129
DB 2817 TTCCACTTCTCCATAGCTTCCATGTTGGCTCAGGACAGACAGAGTCTTCAATGACTGACAA 2876
QY 130 AsnGlyCysCysValThrLysGluGlyArgGlyAla----- 141
```

```
DB 2877 GAGGTGGCCAGCCTCTTAAAAAGGCCAAAGCTCAGCTCTCAGAGATTGAGAGAGTAA 2936
QY 142 -----SerAlaGlyGluGlyLys-GluGlnLysGlnTh 152
DB 2937 GACTCTTAAACAAACCGACAGCCCAAGGACAGAGGTCAAGAAAGTACTCATCAGAGAC 2996
QY 152 rAsnIleGluSerMetLysMetGluGlySerArgGlyArgLeuArgGlyGlyLeuGlyTr 172
DB 2997 CTCTGTGCGAGGACCCCGGATTAACATGTCTGCAGAGAGCAGCTGTGGCCCTTGCCCG 3056
QY 172 pGluSerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTy 192
DB 3057 AAAACGAGCTGTGTTCTCTGATGACATGCCACCCCTGAGT----- 3096
QY 192 rTyrlleSerLysArgLysArgAspGluTrpLeuAlaArgTrpLysArgGluAlaGluLys 212
DB 3097 -----GCCTTACCATGG-----GAGA 3113
QY 212 sLysAlaLysValIleAlaValMetAsnAlaValGluGluAsnGlnAlaSerGlyGluSe 232
DB 3114 ACGAGAAAGATTTTGTCTTCATGGGAATGATCAGAGTCATCAATTCCTGGCTCA-- 3171
QY 232 rGlnLysValGluGluAlaSerProProAlaValGlnGlnProThrAspProAlaSerPr 252
DB 3172 -----GAAGATGCTGACCTCTTCTGCT-----CCACCCATCAACCAATTAACCC 3215
QY 252 oThrValAlaThrThrProGluProValGlyGlyAspAlaGlyAspLysAsnAlaThrLys 272
DB 3216 TCTC-----ACTAGAAA 3227
QY 272 sAlaAlaAspAspGluProGluTrpGluAspGlyArgGlyPheGlyIleGlyGluLeuVa 292
DB 3228 CAAGGCACCCAGGACCTCCAGTAAAGAAAGACGTCG-ATC-----GAGCGGT 3277
QY 292 lTrpGlyLysLeuArgGlyPheSerTrpTrpProGlyArgIleValSerTrpTrpMetTh 312
DB 3278 GTGGCAGTGTCCCG-----CTGCCAGGTGCTGAGGACTGTGGTGTGTGTA 3325
QY 312 rGlyArgSerArgAlaAlaGluGlyThrArgTrpValMetTrpPheGlyAspGlyLysPh 332
DB 3326 CTAATTGCTTAGATAAGCCAA-----GTTGTGTGTCGCATATATAAGA 3370
QY 332 eSerValValCysValGluLysLeu-----MetProLeuSerSerPheCysSerAlaPh 350
DB 3371 AGCAGTCTCGACAGATGAGAAATGTTCAGATCTACATGATGCTTCCAAAGCCTACC 3430
QY 350 eHisGlnAlaThrTyAsnLysGlnProMetTyArgLysAlaIleTyArgLysValLeuG1 370
DB 3431 TGCAGAGCAGCTAAAGCTGTGAAAA-AGAAAGAGAAAAAGTCTAAGACCACTAGTAAAG 3489
QY 370 nValAlaSerSerArgAla----- 376
DB 3490 AAAGACAGCAAGAGAGAGAGAGTGTGTGAGAAACGTTGTGAGACTCTAGTCAGAAACCTACC 3549
QY 377 -----GlyLysLeuPheProAlaCysHisAspSerAspGluSerAspSerGlyLys 393
DB 3550 CCATCAGCAAGAGAGATCTCTGCCCAAAG-AAAAGCAGTAGTAGGCTCTCCACAGAAA 3608
QY 393 sAlaValGluValGlnAsnLysGlnMetIleGluTrpAlaLeuGlyGlyPheGlnProSe 413
DB 3609 CCCCCTCGAGGAAAGAGTGAAGAA-----GGGAATGCTCGGCCCC 3650
QY 413 rGlyProLysGlyLeuGlu----- 419
DB 3651 TGGGCTTGAATCCAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3710
QY 420 -----ProProGluGluGluLysAsn-----ProTyLysGluVa 431
DB 3711 CCAGCCAGCAGCTGGTCTATCCCGCTTCAGCCACTTACAGGACCGCCCAAGAAAGAGT 3770
QY 431 lTyThrAspMetTrpValGluProGluAlaAlaAlaTyAlaProProProAlaLys 451
DB 3771 TCCAAAAACCATCTCTAGTGGAGCCCAAGAAAAAG-----CAGCCTCCACCACAGAGATC 3824
```

```

Qy 451 slyspProArlySer-----ThrtHgl 459
      |||
      |||
      |||
Db 3825 AGGTCCAGAGAGAGCAAGAAAAAGTGGCTCCCGCCCAAGTATCCCTGTAAAAACA 3884
Qy 459 uLyPProLyVallySglu----- 465
      |||
      |||
      |||
Db 3885 AAAACCAAAAGAAAAAGAAACCACTCCGGTCATTAAGCAGAGAAATGACAGCACTTT 3944
Qy 466 ----lLeilaSpLlyrThrtArg----- 472
      |||
      |||
      |||
Db 3945 GAACATCTCTGACCTCTCTCAATGGCAATGTTCTTAAGCAAAAAATTCAGCAGATGG 4004
Qy 473 ----GluArGleuValTyrGluValArgGlnLyseCyse-----ArgAsnileGluasp-- 488
      |||
      |||
      |||
Db 4005 AGTCACACAGGATCAGAGTGGACCTTAAGAGAGATTGTGAACAGAAAAATGTGGAGAT 4064
C 489 -----lLeCySileSerCy 493
      |||
      |||
      |||
Db 4065 GCGAGGCTTAGAAATCTTGACTTCTCTATTAACAACCGAGGCTGTTCTTCTCTG 4124
Qy 493 sGlySerLeuAsnValThrLeuGlnHisProLeuPheileGlyGlyMetCysGlnAsnCy 513
      |||
      |||
      |||
Db 4125 TGCCAGTAGTGG-----CATGTAGAGTTGTG-----TATTCCAAGTCTG 4166
Qy 513 s-----LysAsnCyPheLeuGluCyseAlaTyrGlnTyrAspAspAspG1 528
      |||
      |||
      |||
Db 4167 TTGAGAGCCCTTCCACAAAGTTTGTTTAGAGAGAACAGAGCCCTCTGAGAGACCACT 4226
Qy 528 yTyrGlnSerTyrCysThrIleCyse-----CysGly----- 538
      |||
      |||
      |||
Db 4227 GGAATAATGGTGTGTGCTGCTGCAAAATCTGTACAGCTTTGGAAAGCAACATCAGGC 4286
Qy 539 -GlyArgGluValLeuMetCysGlyAsnAsnAsnCyseYargCysePheCyseValGluCy 558
      |||
      |||
      |||
Db 4287 TACAAGACAGCTGCTGAGGT-----AATTAAGTCCGAAACAGCTATCACTCAGAGT 4340
Qy 558 sValAspLeuLeuValGlyProGly-----AlaAlaGlnAlaIleLysGluAspProTtr 577
      |||
      |||
      |||
Db 4341 C-----CTGGACCAAACTACCCCAACCAACCAAGAAAGAAAGTCTG 4388
Qy 577 PAsnCyseTyrMetCysGlyHisLysGlyThrTyrGlyLeuLeuArgArgGluAspTtr 597
      |||
      |||
      |||
Db 4389 GATCGTATCCAAAGTGTGTTCCTGTGAAGAGCTGTGATCCCAACTCCAGCAAAAGGTG 4448
C 597 pProSerArgLeuGlnMetPhePheAlaAsnHisAsp 610
      |||
      |||
      |||
Db 4449 GGATGCACAGTGTCTCATGATTCTCTCACTGTGTCATGAT 4488

```

```

FILING DATE: September 15, 1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Monaco, Daniel A.
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 8321-8
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-8383
TELEFAX: (215) 568-5549
TELEX: No. 5734039e
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 14255 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-306-691B-55

Alignment Scores:
Pred. No.: 0.0195 Length: 14255
Score: 158.50 Matches: 148
Percent Similarity: 32.17% Conservative: 101
Best Local Similarity: 19.12% Mismatches: 275
Query Match: 3.23% Indels: 254
DB: 1 Gaps: 33

US-09-720-086-5 (1-908) x US-08-306-691B-55 (1-14255)
Qy 10 SerSerSerSerLeuGlnArgGluAspArgLyseGluGlyGlnGlnGlnGluAsn 29
      |||
      |||
      |||
Db 2404 GCTACACAGAGCGCTGGAAGAGCAAGTAGAGAGAGACCGGAGAGAGAAAAAGAG 2463
Qy 30 ArgGlyLyseGlnArgGlnGln-----ProSerAla 40
      |||
      |||
      |||
Db 2464 AATTAAGCGGAGTCAAGAAAGAAAGAAAGAAAGGATCAGAAATTCAGATGATTCT 2523
Qy 41 ThrAlaArgLysValGlyArgProGlyArgLyseArg-----LysHisProProValGluSer 59
      |||
      |||
      |||
Db 2524 GCTTGTATCTCTGGGTAGAGGTTTCCAAAGAGAGAGTGTGTGTAAGATGTTGCCACT 2583
Qy 60 SerAspThrProLysAspProAlaValThrThrLysSerGlnProMetAlaGlnAspSer 79
      |||
      |||
      |||
Db 2584 TCATCTTCTGCCAAA-----AAAGCAACAGCGCGGAAGAGTTCATCATCATGATTCT 2637
Qy 80 GlyPro-----SerAspLeuLeuProAsnGlyAspLeuGluLys 92
      |||
      |||
      |||
Db 2638 GCGACTGATATTACTTCTGTGACTTGTGGGATFACAACAGCTGT-CAAAACCAAAATACT 2696
Qy 93 ArgSerGlnProGlnProGlnGlnGlySerProAlaIleGlyGlnLysGlyGlyAlaPro 112
      |||
      |||
      |||
Db 2697 TATTAAGAAAGAGAGAAATCTGAAAAAACCAACTTGACCTCGGCCCAACTGCCCC 2756
Qy 113 AlaGluGlyGlnGlyThrGlnThrProPro-----GluAlaSerArgAlaValGln 129
      |||
      |||
      |||
Db 2757 ATCCCTGAGAGAGAGAAACCTCTGCTTTCACATCTTTCATCTAGCACTGTTAAACA 2816
Qy 123 ----- 141
      |||
      |||
      |||
Db 2817 TTCACATCTCTCCATAGAGCTCCATGTTGGCTCAGACAGACAAAGCTTCCAAATGACTGACAA 2876
Qy 130 AsnGlyCyseCyseValThrLysGlnGlyArgGlyAla----- 141
      |||
      |||
      |||
Db 2877 GAGGTTGCGCAGCTCTCTAAAGGCCAAAGCTCACTTGCAAGATTGAAAGAGTAA 2936
Qy 142 -----SerAlaGlyGlnGlyLys-GlnGlnLysGlnTh 152
      |||
      |||
      |||
Db 2937 GAGCTTAAACAAACCGACACGCCCAAGCACAGAGGTCAAGAAAGTACATCATAGACAC 2996
Qy 152 rAsnileGluSerMetLysMetGlnGlySerArgGlyArgLeuArgGlyGlyLeuGlyTtr 172
      |||
      |||
      |||

```

Db 2997 CTCTGTCGAGGACCCCGGATTAAACATGTTCTGCAGAGAGACGAGCTGTTGGCCCTGGCCG 3056  
Qy 172 pGluSerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTy 192  
Db 3057 AAAACAGAGCTGTGTTCTCTGATGACATGCCACCCCTGAGT- 3096  
Qy 192 rTyrIleSerLysArgLysArgAspGluTrpLeuAlaArgTrpLysArgLysGluAlaGluLy 212  
Db 3097 - - - - -GCCTTACCATGG- - - - -GAAGA 3113  
Qy 212 sLysAlaLysValIleAlaValMetAsnAlaValGluGlnAlaSerGlyGluSe 232  
Db 3114 ACAGAAAGAGATTGCTCTTCATGGGAATGATGACAGCTCATCAATTGCTGGCTCA- 3171  
Qy 232 rGlnLysValGluGlnAlaSerProProAlaValGlnGlnProThrAspProAlaSerPr 252  
Db 3172 - - - - -GAAGATGCTGAACCTCTTGCT- - - - -CCACCATCAACCACTTAACCC 3215  
Qy 252 oThrValAlaThrThrProGluProValGlyGlyAspAlaGlyAspLysAsnAlaThrLy 272  
Db 3216 TGTC- - - - -ACTAGAAA 3227  
Qy 272 sAlaAlaAspAspGluProGluTrpGluAspGlyArgGlyPheGlyIleGlyGluLeuVa 292  
Db 3228 CAAGGCACCCCGAGACCTCCAGTAAAGAAAGGACGTCG-ATC- - - - -GAGGCGGT 3277  
Qy 292 lTTPGlyLysLeuArgGlyPheSerTrpTrpProGlyArgIleValSerTrpTrpMetTh 312  
Db 3278 GTGGGCGAGTGTCCCG- - - - -CTGCCAGGTGCTGAGGAGCTGTGGTGTGTA 3325  
Qy 312 rGlyArgSerArgAlaAlaGluGlyThrArgTrpValMetTrpPheGlyAspGlyLysPh 332  
Db 3326 CTAATTGCTTAGATAAGCCCA- - - - -GTTTGGTGGTCCCAATATAAGA 3370  
Qy 332 eSerValValCyseValGluLysLeu- - - - -MetProLeuSerSerPheCyseSerAlaPh 350  
Db 3371 AGCAGTGTGCAAGATGAGAAATGTCAGAATCTACAAATGGATGCTTCCAAAGCCTACC 3430  
Qy 350 eHisGlnAlaThrThrAsnLysGlnProMetTyrArgLysAlaIleTyrGluValLeuGl 370  
Db 3431 TGCAGAGAGCACTAAAGCTGTGAAA-AGAAAGAGAAAGTCTAAGACCAAGTGAAGAG 3489  
Qy 370 nValAlaSerSerArgAla- 376  
Db 3490 AAAGACAGCAAGAGAGAGCAGTGTGTGAAGAACGTGTGAGCTCTAGTCAGAAACCTTACC 3549  
Qy 377 - - - - -GlyLysLeuPheProAlaCyseHisAspSerAspGluSerAspSerGlyLy 393  
Db 3550 CCATCAGCAAGAGAGATCCTGCCCCAAAG-AAAAGCAGTAGTAGGCTCTCCACAGAAA 3608  
Qy 393 sAlaValGluValGlnAsnLysGlnMetIleGluTrpAlaLeuGlyGlyPheGlnProSe 413  
Db 3609 GCCCGTCGAGGAAAGAGTGAAGAA- - - - -GGGAATGTCGCGCCCC 3650  
Qy 413 rGlyProLysGlyLeuGlu- 419  
Db 3651 TGGGCTGAATCCAAACAGGCCACCACTCCAGCTTCCAGGAAGTCAAGCAAGCAGGTCTC 3710  
Qy 420 - - - - -ProProGluGluLysAsn- - - - -ProTyrLysGluVa 431  
Db 3711 CCAGCCAGCACTGGTCATCCCGCTCAGCCACCTACTACAGGACCGCCCAAGAAAGAGT 3770  
Qy 431 lTyrThrAspMetTrpValGluProGluAlaAlaIleTyrAlaProProProProAlaLy 451  
Db 3771 TCCCAAAACCACTCTCTAGTGAGCCCAAGAAAAG- - - - -CAGCCTCCACCCAGCAATC 3824  
Qy 451 sLysProArgLysSer- 459  
Db 3825 AGGTCAGAGAGCAGCAACAGAAAAAGTGGCTGCCGCCCAAGTATCCTGTGTAACA 3884  
Qy 459 uLysProLysValLysGlu- 465  
Db 3885 AAAACCAAAAGAAAGGAAACCACTCCGTGTCATTAAGCAGGAGATGACGAGCACTTT 3944

Qy 466 - - - - -IleIleAspGluArgThrArg- 472  
Db 3945 GAACATCTCTCAGCACTCTCCAATGCAATGTTCTTAAGCAAAAAAATTCACGACAGATGG 4004  
Qy 473 - - - - -GluArgLeuValTyrGluValArgGlnLysCyse- - - - -ArgAsnIleGluAsp- 488  
Db 4005 AGTCCACAGGATCAGATGGAGCTTTAAGGAGGATTTGTAAGCAGAAATTTGTGGGAGAT 4064  
Qy 489 - 493  
Db 4065 GGGAGGCTTAGAATCTTGACTTCTGTTCTTAACACCCAGGGTGGTTGCTTTCTCTG 4124  
Qy 493 sGlySerLeuAsnValThrLeuGluHisProLeuPheIleGlyGlyMetCyseGlnAsnCy 513  
Db 4125 TGCAGTAGTGGG- - - - -CATGTAGAGTTTGTG- - - - -TATTGCCAAGTCTG 4166  
Qy 513 s- - - - -LysAsnCysePheLeuGluCyseAlaTyrGlnTyrAspAspG1 528  
Db 4167 TTGTGAGCCCTTCCACAAGTTTGTGTTTAGAGGAGAACGAGCGCCCTCTGGAGGACCAGCT 4226  
Qy 528 yTyrGlnSerTyrCyseThrIleCyse- - - - -CyseGly- - - - - - - - - - - 538  
Db 4227 GGAATAATTGGTGTGCTGCTGCAAAATCTGTCACTGTTTGTGAAGGCAACATCAGGC 4286  
Qy 539 -GlyArgGluValLeuMetCyseGlyAsnAsnAsnCyseArgCysePheCyseValGluCy 558  
Db 4287 TACAAAGCAGCTGTGGAGTGT- - - - -AATAAGTGCCGAACAGCTATCACCCTGAGTG 4340  
Qy 558 sValAspLeuValGlyProGly- - - - -AlaAlaGlnAlaIleLysGluAspProTr 577  
Db 4341 C- - - - -CTGGACCAAACTACCCCAACCAAGGAGTGTGATCCACAACTCCAGCAAGGGTG 4388  
Qy 577 pAsnCyseTyrMetCyseGlyHisLysGlyThrTyrGlyLeuLeuArgArgGluAspTr 597  
Db 4389 GATCTGTACCAAGTGTGTCGCTGAAGAGCTGTGATCCACAACTCCAGCAAGGGTG 4448  
Qy 597 pProSerArgLeuGlnMetPhePheAlaAsnHisAsp 610  
Db 4449 GGATGCACAGTGGTCTCATGATTCTCACTGTGTCATGAT 4488

RESULT 12  
US-08-545-860D-1  
; Sequence 1, Application US/08545860D  
; Patent No. 6040140  
; GENERAL INFORMATION:  
; APPLICANT: Croce, Carlo  
; APPLICANT: Canaan, Eli  
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods  
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias  
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region  
; NUMBER OF SEQUENCES: 94  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &  
; ADDRESS: No. 604014018  
; STREET: One Liberty Place, 46th floor  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/545,860D  
; FILING DATE: 07-MAR-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/04496  
; FILING DATE: 22-APR-1994  
; PRIOR APPLICATION DATA:



```
QY 420 -----ProGluGluGluLysAsn-----ProTyLysGluVa 431
Db 3711 CCAGCCAGCACTGTCATCCCGCTCAGCCAGCACTTACAGGACCGCCCAAGAAAGAGT 3770
QY 431 lTyThrAspMetTrpValGluProGluAlaAlaAlaTyAlaProProProProAlaLy 451
Db 3771 TCCCAAAACCACTCTAGTAGGAGCCCAAGAAAG-----CAGCTCCACCAAGATC 3824
QY 451 sLysProArgLysSer-----ThThrGl 459
Db 3825 AGTCCAGAGCAGAGCAACACAGAAAAAGTGGCTCCCGGCCCAAGTATCCCTGTAAACA 3884
QY 459 uLysProLysValLysGlu----- 465
Db 3885 AAAACCAAAAGAAAGAAAAAACCAACCTCCCGTCAATAAGCAGGAGAAATGCAGGACATTT 3944
QY 466 ----IleIleAspGluArgThrArg----- 472
Db 3945 GAACATCTCTCAGCACTCTCTCCAATGGCAATAGTTTAAAGCAAAAAATCCAGCAGATGG 4004
QY 473 ----GluArgLysValTyGluValArgGlnLysCys-----ArgAsnIleGluAsp-- 488
Db 4005 AGTCCACAGATCAGAGTGGACTTTTAAAGGAGGATTTGTGAAGCAGAAAAATGTGTGGAGAT 4064
QY 489 -----IleCysIleSerCy 493
Db 4065 GGGAGGCTTAGAATCTTGACTTCTGTTCTTAAACACCCAGGCTGTTGCTTCTCTG 4124
QY 493 sGlySerLysAsnValThrLeuGluHisProLeuPheIleGlyMetCysGlnAsnCy 513
Db 4125 TCCAGTAGTGGG-----CATGTAGAGTTTGTG-----TATTCGCAAGTCTG 4166
QY 513 s-----LysAsnCyPheLeuGluCysAlaTyArgLntTyAspAspAspGl 528
Db 4167 TTGTAGCCCTTCCCAAGTTTGTGTAGAGGAGAGAGCGGCCCTCTCGGAGGACCACT 4226
QY 528 YTyGlnSerTyxCysThrIleCys-----CysGly----- 538
Db 4227 GGAATAATTGGTGTGTGCTGCTGTCGAAATTTCTGCACGTTTGTGGAAGCAACATCAGGC 4286
QY 539 -GlyArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCy 558
Db 4287 TACAAGCAGCTCTGGAGTGT-----ATAAGTGCAGAAACAGCTATCACCTGAGTG 4340
QY 558 sValAspLeuLeuValGlyProGly---AlaAlaGlnAlaAlaIleLysGluAspProTr 577
Db 4341 C-----CTGGACCAAACTACCCCAACCAACCCCAAGAAAGAAAGTCTG 4388
QY 577 pAsnTyMetCysGlyHisLysGlyThrTyGlyLeuLeuArgArgArgGluAspTr 597
Db 4389 GATCTGTACCAAGTGTGTGCTGTAAAGAGCTGTGGATCCACAACCTCCAGGCAAGGTTG 4448
QY 597 pProSerArgLeuGlnMetPhePheAlaAsnAsnHisAsp 610
Db 4449 GGATGCACAGTGGTCTCATGATTTCTCAGCTGTGTCATGAT 4488

RESULT 13
PCT-US94-04496-1
; Sequence 1, Application PC/TUS9404496
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canani, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESSEE: Norris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
```

```
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04496
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Deluca Esq., Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-1242
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 14255
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ANTI-SENSE: No
PCT-US94-04496-1

Alignment Scores:
Pred. No.: 0.0195 Length: 14255
Score: 158.50 Matches: 146
Percent Similarity: 32.17% Conservative: 101
Best Local Similarity: 19.12% Mismatches: 275
Query Match: 5.23% Indels: 254
DB: 5 Gaps: 33

US-09-720-086-5 (1-908) x PCT-US94-04496-1 (1-14255)
QY 10 SerSerSerSerLeuGluArgGluAspAspArgLysGlyGluGluGluGluGluAsn 29
Db 2404 GCTCAAGAGCGTGGAGGAGCAAGTAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2463
QY 30 ArgGlyLysGluGluArgGlnGlu-----ProSerAla 40
Db 2464 AATAAGCGGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2523
QY 41 ThrAlaArgLysValGlyArgProGlyArgLysArg---LysHisProProValGluSer 59
Db 2524 GCTTTGTATCTGTTGGTAGGTTTCCAAAGAGAGAGTGTGTGTGAAGATGTTGCCACT 2583
QY 60 SerAspThrProLysAspProAlaValThrThrLysSerGlnProMetAlaGlnAspSer 79
Db 2584 TCATCTTCGCCAAA-----AAAGCAGAGCGCGGAGAGAGTCTTCATCATGATTTCT 2637
QY 80 GlyPro-----SerAspLeuLeuProAsnGlyAspLeuGluLys 92
Db 2638 GGGAGCTGATATTACTTCTGTGACTCTTTGGGATACACAGAGCTGT-CAAAACCAATACT 2696
QY 93 ArgSerGluProGlnProGluGluGlySerProAlaAlaGlyGlnLysGlyGlyAlaPro 112
Db 2697 TATAAAGAGAGGAGAGAGAAATCTGGAAAAAACCAACTTGGACCTCGGCCCAACTGCC 2756
QY 113 AlaGluGlyGluGlyThrGluThrProPro----- 122
Db 2757 ATCCCTGGAGAGAGAGAGAAACCCCTCTGCCTTCCACTCTCATCTAGCACTGTAAACA 2816
QY 123 -----GluAlaSerArgAlaValGlu 129
Db 2817 TTCCACTTCTCCATAGCTCCATGTTGGCTCAGGAGAGAGAGTTCATGACTGACAA 2876
QY 130 AsnGlyCysCysValThrLysGluGlyArgGlyAla----- 141
Db 2877 GAGGTTGACAGCTCTCTAAAAAGGCAAGCTCAGCTCTGCAAGATTGAGAGAGATAA 2936
QY 142 -----SerAlaGlyGlyLys-GluGlnLysGlnTh 152
```



ATTORNEY/AGENT INFORMATION:

```

1 NAME: PARKER, DAVID L.
2
3 REGISTRATION NUMBER: 32 165
4 REFERENCE/DOCKET NUMBER: ARCD.072/PAR
5 INFORMATION:
6 TELECOMMUNICATION INFORMATION:
7 TELEPHONE: (512) 320-7200
8 TELEFAX: (512) 474-7577
9
10 INFORMATION FOR SEQ ID NO: 4:
11 SEQUENCE CHARACTERISTICS:
12 LENGTH: 4201 base pairs
13 TYPE: NUCLEIC ACID
14 STRANDEDNESS: single
15 TOPOLOGY: linear
16
17 MOLECULE TYPE: DNA (genomic)
18
19 US-08-080-255-4

```

Alignment Scores:		
Pred. No.:	0.00427	Length:
Score:	157.00	Matches:
Percent Similarity:	34.7%	Conservative:
Best Local Similarity:	20.1%	Mismatches:
Query Match:	3.20%	Indels:
DB:	1	Gaps:
		4201
		148
		105
		274
		211
		32

US-09-720-086-5 (1-908) x US-08-080-255-4 (1-4201)

Qy	10	SerSerSerSerLeuGluArgGluuPaspAArgLysGluGlyGluGlnGluGluuSen	29
Db	257	CTCAGCAAGAGCGCTGGAGAGGACAGAGTAGTACAGAGACAGACC	316
Qy	30	ArgGlyLysGluGluArgGlnGlu	40
Db	317	AATAAAGCGGGAGTCAAAGGAAAGAGAAAGAAAGGAGATCAGAAATTCAGAGTAGTGTTCT	376
Qy	41	ThrAlaArgLysValGlyArgProGlyArgLysArg--LysHisProValGluSer	59
Db	377	GCTTTGTATCCCTGTGGTAGGGTTTCCAAGAGAAAGGTGTGGTGAAGATGTTGGCCACT	436
Qy	60	SerAspThrProLysAspProAlaValThrThryLysSerGlnProMetAlaGluAspSer	79
Db	437	TCATCTTCTGCCAAA-----AAGCAACAGCGCGGAAGAGTCTTCATCATCATGATTCT	490
Qy	80	GlyPro-----SerAspLeuLeuProAsnGlyAspLeuGluLys	92
Db	491	GGGACTGATATTACTTCTGTGACTCTTGGGGATACACAGCTGT-CAAAACCAAAAATACT	549
Qy	93	ArgSerGluProGlnProGluGluGlySerProAlaAlaGlyGlnGlyGlyValaPro	112
Db	550	TATAAAGAAAGGAGGAGGAAATCTGGAAAAACCAACTTGGACTCGCCCAACTGCCCC	609
Qy	113	AlaGluGlyGlyGlyThrGluThrProPro-----	122
Db	610	ATCCCTCGGAAGGAGAAACCCCTCTGCCCTTCTCCACTCTTCATCTAGCACTGTGTAACA	669
Qy	123	-----GluAlaSerArgAlaValaGlu	129
Db	670	TTCCACTCTCCTCATAGGCTCCATGTGGCTCAGGACAGCAAGCTTCCAATGACTGACAA	729
Qy	130	AsnGlyCysValThrLysGluGlyArgGlyAla-----	141
Db	730	GAGGGTTTCCAGGCTCCTAAAAAAGGCCAAAGCTCAGCTCTGCAAGATTGAGAAGAGATAA	789
Qy	142	-----SerAlaGlyGluGlyLys-GluGlnLysGlnTh	152
Db	790	GAGTCTTAAACAAACCGACACCGCCAAAGCAGGGTCAAGAAAGTGACTCATCAGAGAC	849
Qy	152	AsnIleGluSerMetLysMetGluGlySerArgGlyValGluArgGlyGlyLeuGlyTr	172
Db	850	CTCTGTGCGAGGACCCCGGATTAAACATGCTGCGAAGAGCAGCTGTGGCCCTTGGCCG	909
Qy	172	pGluSerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTy	192
Db	910	AAAACGAGCTGTCTTCTCGATGATCATGCTCCACCTCAGT-----	949



QY 502 -----HisProleu-----PheIleGlyMetCysGlnAsnCy 513  
Db 1849 TCACAGATGAGTCCACAGAGTGAAGTTGTGATTGCCAAGTCTG 1908  
QY 513 s-----LysAsnCyPheLeuGluCysAlaTyrGlnTyrAspAspAspG 528  
Db 1909 TTGAGACCTTCCACAAAGTTTGTATTAGAGAGAACGAGGCCCTTGGAGGACAGCT 1968  
QY 528 YTyGlnSerTyrCysThrIleCys-----CysGly----- 538  
Db 1969 GGAATAATGGTGTGTGCTGCTGCAATTCTGTACAGTTTGTGGAAGGCAATCAGGC 2028  
QY 539 -GlyArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCy 558  
Db 2029 TACAAACACCTGCTGAGTGT-----AATAGTGCAGAAACAGCTATCACCCTGAGTG 2082  
QY 558 sValAspLeuLeuValGlyProGly---AlaAlaGlnAlaAlaIleLeuGluAspProTyr 577  
Db 2083 C-----CTGGAGCCAAACTACCCACCAAAACCCCAAGAAAGAAAGTCTG 2130  
QY 577 PAsnCysTyrMetCysGlyHisIleGlyThrTyrGlyLeuLeuArgArgArgGluAspTyr 597  
Db 2131 GATCTGATACCAAGTGTGTGCTGTGAAGAGCTGTGATCCCAACTCCAGGCAAGGCTG 2190  
QY 597 PProSerArgLeuGlnMetPhePheAlaAsnAsnHisAsp 610  
Db 2191 GGATGCACAGTGTCTCATGATTCTCAGCTGTGTATGAT 2230

## RESULT 15

US-08-465-713-4  
Sequence 4, Application US/08465713  
Patent No. 6121419  
GENERAL INFORMATION:  
APPLICANT: Rowley, Janet D.  
APPLICANT: Diaz, Manuel O.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P. O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,713  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/080,255  
FILING DATE: 17 JUNE 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, David L.  
REGISTRATION NUMBER: 32,165  
REFERENCE/DOCKET NUMBER: ARCD:072/PAR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 330-7200  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4201 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-465-713-4

Alignment Scores:  
Pred. No.: 0.00427 Length: 4201  
Score: 157.00 Matches: 148  
Percent Similarity: 34.47% Conservative: 105  
Best Local Similarity: 20.16% Mismatches: 274  
Query Match: 3.20% Indels: 211  
DB: 3 Gaps: 32  
US-09-720-086-5 (1-908) x US-08-465-713-4 (1-4201)  
QY 10 SerSerSerSerLeuGlnValGluAspAspArgIleGluGluGluGluGluAsn 29  
Db 257 GCTGACAGAGCGCTGGAGAGGACAAGAGTAGAGAGAGACCGGAGAGAGAGAGAG 316  
QY 30 ArgGlyLeuGluValGluGluGluGluGluGluGluGluGluGluGluGluGlu 40  
Db 317 AATAGCGGAGTCAAG 376  
QY 41 ThrAlaArgIleValGlyArgProGlyArgIleArgIleArgIleArgIleArgIle 59  
Db 377 GCTTTGATCTGTGGGTAGAGTTTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAG 436  
QY 60 SerAspThrProIleAspProAlaValThrThrIleSerGlnProMetAlaGlnAspSer 79  
Db 437 TCATCTTCTGCCAAA-----AAGCAACAGGCGGAGAGAGAGAGAGAGAGAG 490  
QY 80 GlyPro-----SerAspLeuLeuProAsnGlyAspLeuGluLeu 92  
Db 491 GGGAGTATATTACTTGTGACTCTTGGGATACACAGCTGT-CAAAACAAATACT 549  
QY 93 ArgSerGluProGlnProGluGluGlySerProAlaAlaGluGluGlyGlyAlaPro 112  
Db 550 TATTAAGAAAGGAG 609  
QY 113 AlaGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 122  
Db 610 ATCCCTGAG 669  
QY 123 -----GluIleSerAlaValAlaGlu 129  
Db 670 TTCCACTCTCTCCATAGCTCCATGTTGGCTCAGGACAGAGAGAGAGAGAGAGAG 729  
QY 130 AsnGlyCysCysValThrIleGluGlyArgGlyAla----- 141  
Db 730 GAGGTTGCCAGCTCTTAATAAAGCCCAAGCTCAGCTGTGCAAGTTGAGAAGACTAA 769  
QY 142 -----SerAlaGlyGluGlyIleGluGluGluGluGluGluGluGluGlu 152  
Db 790 GAGTCTTAACAAACGAGCCAGCCCAAGCAGAGGTCAGAGAAAGTACTCATCAGAGAC 849  
QY 152 rAnIleGluSerMetIleMetIleGluGlySerArgGlyArgLeuArgGlyGlyLeuGlyTyr 172  
Db 850 CTCTGTGCGAGAGACCCCGGATTAACATGTCTGACAGAGAGAGAGAGAGAGAGAG 909  
QY 172 pGluSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyr 192  
Db 910 AAAACAGCTGTGTTCTTCTATGATGATGCCACCTGAGT----- 949  
QY 192 rTyIleSerLeuArgIleArgIleArgIleArgIleArgIleArgIleArgIleArgIle 212  
Db 950 -----GCCTTACCATGG-----GAAGA 966  
QY 212 sIleValAlaValAlaValAlaValAlaValAlaValAlaValAlaValAlaValAla 232  
Db 967 ACAGAGAAAGATTTTGTCTTCATGGGAGATGATGACCAATCTCATATTCCTGCTCA-- 1024  
QY 232 rGlnIleValGluGluAlaSerProProAlaValGlnGlnProThrAspProAlaSerPr 252  
Db 1025 -----GAAGATGCTGAACCTCTGTCT-----CCACCATCAACCAATTAAAC 1068  
QY 252 oThrValAlaThrThrProGluProValGlyIleAspAlaGlyAspIleAsnAlaThrIle 272

Db 1069 TGTC-----ACTAGAAA 1080  
Qy 272 sAlaAlaAspGluProGluTyrGluAspGlyArgGlyPheGlyIleGlyGluLeuVa 292  
Db 1081 CAAGGCACCCAGCAACCTCCAGTAAGAAAGAGCGTCG-ATC-----GAGGCGGT 1130  
Qy 292 lTTPGlyLysLeuArgGlyPheSerTTPTPProGlyArgIleValSerTTPTPMerTh 312  
Db 1131 GTGGCAGTGTCCGG-----CTGCCAGGTGCCTGAGGACTGTGTGTGTGTA 1178  
Qy 312 rGlyArgSerArgAlaAlaGluGlyThrArgTPValMetTTPPheGlyAspGlyLysPhe 332  
Db 1179 CTAATTCCTTAGATAAGCCAA-----GTTTGGTGTCCCAATATAAAGA 1223  
Qy 332 eSerValValCysValGluLysLeu-----MetProLeuSerSerPheCysSerAl 349  
Db 1224 AGCAGTCTCCAAGATGAGAAATGTCAAGAATCTACTACAATCGATGCGCTTCCAAAGCCT 1283  
Qy 349 aPheHisGlnAlaThrTyrAsnLysGlnProMetTyrArgLysAlaIleTyrGluValLe 369  
Db 1284 ACCTGCAGAGCAAGCTAAAGCTGTGAAAA-AGAAAGAGAAAGTCTAAGACCACTGAA 1342  
Qy 369 uGlnValAlaSerSerArgAla-----376  
Db 1343 AAGAAACACAGCAAGACAGACGAGTGTGTGAAGAACGTGTGGACTCTAGTCAGAAACCT 1402  
Qy 377 -----GlyLysLeuPheProAlaCysHisAspSerAspGluSerAspSerGl 392  
Db 1403 ACCCCATCAGCAAGACGAGGATCCTGCCCAAAG-AAAAGCAGTAGTGAGCCTCTCCACG 1461  
Qy 392 yLysAlaValGluValGlnAsnLysGlnMetIleGluTTPAlaLeuGlyGlyPheGlnPr 412  
Db 1462 AAAGCCGCTCGAGAAAGAGTGAAGAA-----GGCAATGTCTCGGC 1503  
Qy 412 oSerGlyPro-----LysGlyLeuGluProProGluGluGlyLysAsnProTyrLysGl 430  
Db 1504 CCCTGGCCCTGAATCCAAACAGGCCACCACTCCAGCTTCAGGAGTCAAGCAAGCAGGT 1563  
Qy 430 uValTyrThrAspMetTTPValGluProGluAlaAlaIleTyrAlaProPro----- 447  
Db 1564 CTCCAGCCAGCACTGGTCTATCCGCCCTCAG-----CCACCTACTACAGG 1608  
Qy 448 -ProProAlaLysProArgLysSerThrThrGluLysProLysValLys----- 464  
Db 1609 ACCGCCAAGAAAGAGTTCCTCCAAACCACTCTCTAGTGAGCCCAAGAAAGACGACCTCC 1668  
Qy 465 -----GluIleIleAspGluArgThrArgLysLeuVal-----Ty 477  
Db 1669 ACCACCAAGATCAGGTCCAGAGCAGAGCAAGCAAGAAAGTGGCTCCCGCCCAAGTAT 1728  
Qy 477 rGluValArgGlnLysCysArgAsnIleGlu-----AspIleCysIleSe 492  
Db 1729 CCCTGTAAAAACAAAACCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1788  
Qy 492 rCysGlySerLeuAsnVal-----ThrLeuGlu----- 501  
Db 1789 TGCAGGCACCTTGAACATCTCCAGCACTCTCTCCAATGGCAATAGTCTTAAGCAAAAAAT 1848  
Qy 502 -----HisProLeu-----PheIleGlyMetCysGlnAsnCy 513  
Db 1849 TCCAGCAGATGGAGTCCAGAGATCAGAGTGGAGCTTTAAGTTTGTGTATTCGCAAGTCTG 1908  
Qy 513 s-----LysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAspGln 528  
Db 1909 TTGTAGCCCTTCCACAAGTTTGTGTAGAGGAGAACGAGCGCCCTCTGGAGGACCACT 1968  
Qy 528 yTyrGlnSerTyrCysThrIleCys-----CysGly----- 538  
Db 1969 GAAAAATTTGT 2028  
Qy 539 -GlyArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCy 558  
Db 2029 TACAAAGCAGCTCTGGAGTGT-----AATAAGTCCGAAACAGCTATCACCCCTGAGTG 2082

Qy 558 sValAspLeuLeuValGlyProGly---AlaAlaGlnAlaAlaIleLysGluAspProTr 577  
Db 2083 C-----CTGGACCAAACTACCCCAACCAACCAAGAAAGAAAGTCTG 2130  
Qy 577 pAsnCysTyrMetCysGlyHisLysGlyThrTyrGlyLeuLeuArgArgGluAspTr 597  
Db 2131 GATCTGTACCAAGTGTGTCTGCTGAAGAGCTGTGGATCCCAACTCCAGGCAAGGGTG 2190  
Qy 597 pProSerArgLeuGlnMetPhePheAlaAsnAsnHisAsp 610  
Db 2191 GGATGCACAGTGGTCTCATGATTCTCACCTGTGTCTATGAT 2230

Search completed: November 25, 2002, 01:45:03  
Job time : 312.239 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2002, 08:18:31 ; Search time 84.7727 Seconds  
(without alignments)  
15175.990 Million cell updates/sec

Title: US-09-720-086-2

Perfect score: 4195  
Sequence: 1 gaattccggcgccgggggtc.....aaaaaaaaaaaaaaaaaaaa 4195

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Sequences: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/prodata/1/ina/5A-COMB.seq: \*  
2: /cgn2\_6/prodata/1/ina/5B-COMB.seq: \*  
3: /cgn2\_6/prodata/1/ina/6A-COMB.seq: \*  
4: /cgn2\_6/prodata/1/ina/6B-COMB.seq: \*  
5: /cgn2\_6/prodata/1/ina/PCTUS-COMB.seq: \*  
6: /cgn2\_6/prodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	548.4	13.1	2077	4	US-09-276-531-47
2	49.4	1.2	7218	1	US-08-232-463-14
3	45.2	1.1	7218	1	US-08-232-463-14
4	43.2	1.0	1332	4	US-09-333-423-1
5	42.8	1.0	812	4	US-09-091-097-7
6	42.4	1.0	1046	1	US-08-361-467B-4
7	42.4	1.0	1046	1	US-08-484-332C-4
8	42.2	1.0	2239	4	US-09-196-390-1
9	42	1.0	1378	4	US-09-149-476-208
10	42	1.0	2323	4	US-09-149-476-24
11	41.8	1.0	271	2	US-08-731-272A-29
12	41.4	1.0	2065	4	US-09-370-473-5
13	41.4	1.0	2091	4	US-09-813-818-1
14	40.8	1.0	270	2	US-08-520-678A-30
15	40.8	1.0	270	2	US-08-897-126-30
16	40.8	1.0	790	4	US-09-461-474-13
17	40.8	1.0	3637	1	US-08-445-440-3
18	40.8	1.0	3637	1	US-08-170-558-3
19	40.8	1.0	3637	3	US-08-447-314-3
20	40.8	1.0	3637	3	US-08-445-461-3
21	40.4	1.0	1578	4	US-09-416-050A-1
22	40.4	1.0	1578	4	US-09-664-800-1
23	40.4	1.0	1578	4	US-09-665-309-1
24	40.4	1.0	1578	4	US-09-661-569-1
25	40.2	1.0	635	1	US-08-455-633A-35
26	40.2	1.0	635	1	US-08-416-336-5
27	40.2	1.0	635	2	US-08-456-460C-35

28	40.2	1.0	635	5	PCT-US94-05354-35	Sequence 35, Appl
29	40.2	1.0	1100	3	US-09-248-335-53	Sequence 53, Appl
30	40.2	1.0	1466	4	US-08-984-919A-10	Sequence 10, Appl
31	40.2	1.0	1466	4	US-08-984-919A-12	Sequence 12, Appl
32	40.2	1.0	1472	4	US-08-781-420-10	Sequence 10, Appl
33	40.2	1.0	1472	4	US-08-781-420-12	Sequence 12, Appl
34	40.2	1.0	1472	4	US-08-874-102-10	Sequence 10, Appl
35	40.2	1.0	1472	4	US-08-874-102-12	Sequence 12, Appl
36	40.2	1.0	1472	4	US-09-006-595A-10	Sequence 10, Appl
37	40.2	1.0	1472	4	US-09-006-595A-12	Sequence 12, Appl
38	40.2	1.0	1493	6	5340934-5	Patent No. 5340934
39	40.2	1.0	1875	4	US-08-984-919A-46	Sequence 46, Appl
40	40.2	1.0	1875	4	US-08-984-919A-48	Sequence 48, Appl
41	40.2	1.0	1881	4	US-08-874-102-46	Sequence 46, Appl
42	40.2	1.0	1881	4	US-08-874-102-48	Sequence 48, Appl
43	40	1.0	2852	3	US-09-027-137-2	Sequence 2, Appl
44	40	1.0	2852	3	US-09-344-441-2	Sequence 2, Appl
45	39.8	0.9	337	2	US-09-032-684-11	Sequence 11, Appl

#### ALIGNMENTS

RESULT 1  
US-09-276-531-47  
Sequence 47, Application US/09276531  
Patent No. 6183968  
GENERAL INFORMATION:  
APPLICANT: Bardman, Olga  
APPLICANT: Lai, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Yue, Henry  
APPLICANT: Reddy, Roopa  
APPLICANT: Guegler, Karl J.  
APPLICANT: Baughn, Marich R.  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276,531  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/079,677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lynn E. Murry, Ph.D.  
REGISTRATION NUMBER: 42,918  
REFERENCE/DOCKET NUMBER: PA-0008 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2077 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: TESTTUT02







Patent No. 5767374  
GENERAL INFORMATION:  
APPLICANT: De Greef, Willy  
APPLICANT: Van Emmelo, John  
APPLICANT: De Oliveira, Dulce B.  
APPLICANT: De Souza, Maria-Helena  
APPLICANT: Van Montagu, Marc  
TITLE OF INVENTION: PLANTS WITH MODIFIED FLOWERS, SEEDS OR  
NUMBER OF SEQUENCES: 13  
TITLE OF INVENTION: EMBRYOS  
CORRESPONDENCE ADDRESS:  
ADDRESSER: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,332C  
FILING DATE: 7-JUNE-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/361,467  
FILING DATE: 22-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/681,492  
FILING DATE: 04-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/EP90/01275  
FILING DATE: 01-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 89 402 224.3  
FILING DATE: 04-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Schulman, Robert M.  
REGISTRATION NUMBER: 31,196  
REFERENCE/DOCKET NUMBER: 010830-093  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
ORINATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1046 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
IMMEDIATE SOURCE:  
CLONE: 3C9  
US-08-484-332C-4

Query Match 1.0%; Score 42.4; DB 1; Length 1046;  
Best Local Similarity 54.4%; Pred. No. 0.069; Mismatches 66; Indels 0; Gaps 0;  
Matches 82; Conservative 0;  
QY 3074 CATGTCGTCATGTTCTCTTACCTGAACCTTAAACCTGAAGTAGTAGATGAGC 3133  
DB 889 CGTAGTTCTTCTGTTCTTCTATTTGTTTCTCTATCAAAACCAACAGTAAATGGA 948  
QY 3134 TTTCTTTACCTCTGAGTTTATCATCTGAGAGTGTGGTAGATACCAAAAAACAA 3193  
DB 949 TTTATAAGTTTCTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1008  
QY 3194 ACAAAAAACAGAAACAAAAACAAAAA 3221  
DB 1009 AAAAAAAAAAAAAAAAAAAAAAAAAA 1036

RESULT 8  
US-09-196-390-1  
Sequence 1, Application US/09196390  
Patent No. 6307125  
GENERAL INFORMATION:  
APPLICANT: Block, Martina  
APPLICANT: Lotz, Horst  
APPLICANT: Lutticke, Stephanie  
APPLICANT: Walther, Lennart  
APPLICANT: Froberg, Claus  
APPLICANT: Koemann, Jens  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING ENZYMES  
TITLE OF INVENTION: FROM WHEAT WHICH ARE INVOLVED IN STARCH  
NUMBER OF SEQUENCES: 9  
TITLE OF INVENTION: SYNTHESIS  
CORRESPONDENCE ADDRESS:  
ADDRESSER: James F. Haley, Jr., c/o Fish & Neave  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (BPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/196,390  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 196 21 588.9  
FILING DATE: 28-MAY-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 196 36 917.7  
FILING DATE: 11-SEP-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP97/02793  
FILING DATE: 28-MAY-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley, Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: AGREVO-9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2239 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Triticum aestivum L.  
STRAIN: cv. Florida  
HAPLOTYPE: ca. 21 d Caryopses  
IMMEDIATE SOURCE:  
LIBRARY: cDNA library in paluescript sk (-)  
CLONE: Tasss  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..2017  
US-09-196-390-1

Query Match 1.0%; Score 42.2; DB 4; Length 2239;  
Best Local Similarity 54.1%; Pred. No. 0.12; Mismatches 73; Indels 0; Gaps 0;  
Matches 86; Conservative 0;  
QY 3073 GCATGTCGTCATGTTCTCTTACCTGAACCTTAAACCTGAAGTAGTAGATGAGC 3132

```
Db 2042 GCCTTCTGTGTATGTTGTTCTTCTTACCTGACAAATATTTGACCTGTGGAGAAATTTT 2101
Oy 3133 CTTTCTTTTACCTCTCAGTGTATCTACCTCAGAGTGATGGCTAAGATACCAAAAAAACA 3192
Db 2102 ATCTTGTCTGCTGTTTTTTTTTTTAAATCAAAAGAGGGGTTTCTCTCCGATTTCTATTAATA 2161
Oy 3193 AACAAAAACAGAAAAACAAAAACAAAAACCTCAACA 3231
Db 2162 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2200

RESULT 9
US-09-149-476-208
; Sequence 208, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; EARLIER FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582

; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,880
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,894
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,911
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,910
; EARLIER FILING DATE: 1997-08-22
```





```

; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,880
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,894
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,911
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,910
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,864
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,631

```

```

; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,845
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,892
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,761
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/047,595
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,599
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,588
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,585
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,586
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,590
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,594
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,589
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,593
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,614
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,578
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,576
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/047,501
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,670
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/056,632
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,664
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,876
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,881
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,909
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,875
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,862
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,887
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,908
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/057,650
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/056,884
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,669
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/049,610
; EARLIER FILING DATE: 1997-06-13
; EARLIER APPLICATION NUMBER: 60/061,060
; EARLIER FILING DATE: 1997-10-02

```

Query Match 1.0%; Score 42; DB 4; Length 2323;

Best Local Similarity 60.5%; Pred. No. 0.14; Matches 69; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```

QY 3118 TAGGTAGTAAGATGCTTTCTTTTACCTCCCTCGAGTTTATCAGAGTGTGCTAA 3177
Db 2186 TATGTACAACTGATGCTTGTCTTATTTTAAATAAATTTATCAGAGTCAAAAAA 2245

```





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 24, 2002, 12:10:37 ; Search time 67.8633 Seconds  
(without alignments)  
3854.741 Million cell updates/sec

Title: US-09-720-086-8  
Perfect score: 4566  
Sequence: 1 MKGDTRLHNGEDDAGREDS.....MSVFVIRHLPLKDYFACE 853

Scoring table:  
BLOSUM62  
Xgapod 10.0 , Xgapext 0.5  
Ygapod 10.0 , Ygapext 0.5  
Fgapod 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=Frame+ p2n.model -DEV=x1h  
-DB=/cgn2\_1/USPFO.spool/US09720086/runat.18112002.092957.29457/app.query.fasta\_1.4252  
-DB-issued Patents NA -OPMT=fastcap -SUFFIX=rni -MINMATCH=0.1 -DOFFC=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=numan40.cdt  
-LIST=45 -DOCCALING=200 -THR SCORE=Pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09720086 @CGN 1 1 108 @runat.18112002.092957.29457 -NCPU=6 -ICPU=3  
-NO XLPXY -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WART TIMEOUT=30 -THREADS=1 -XGAPOD=10 -XGAPEXT=0.5 -Fgapod=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/p10data/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/p10data/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/p10data/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/p10data/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/p10data/1/ina/PCtus.COMB.seq:\*  
6: /cgn2\_6/p10data/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	903	19.8	2077	4	US-09-276-531-47 Sequence 47, Appl
2	148.5	3.3	6328	4	US-08-913-832A-1 Sequence 1, Appl
3	148.5	3.3	6328	4	US-09-249-181A-1 Sequence 1, Appl
4	146	3.2	2355	4	US-08-913-159-12 Sequence 12, Appl
5	144.5	3.2	3522	4	US-09-484-970B-8 Sequence 8, Appl
6	137.5	3.0	4481	4	US-09-041-886-18 Sequence 18, Appl
7	137.5	3.0	36778	4	US-09-105-537-5 Sequence 5, Appl
8	137.5	3.0	38506	3	US-09-320-878-19 Sequence 19, Appl
9	133	2.9	35524	3	US-08-923-137-1 Sequence 1, Appl
10	130.5	2.9	2376	2	US-08-760-745-4 Sequence 4, Appl
11	129.5	2.8	80161	3	US-09-036-987A-1 Sequence 1, Appl
12	129.5	2.8	80161	4	US-09-370-700-1 Sequence 1, Appl

13	126	2.8	8580	4	US-09-491-772-1 Sequence 1, Appl
14	125.5	2.7	2874	4	US-09-179-558-54 Sequence 54, Appl
15	125.5	2.7	17612	3	US-08-911-853-29 Sequence 29, Appl
16	125.5	2.7	17612	4	US-09-479-409-29 Sequence 29, Appl
17	125.5	2.7	17612	4	US-09-479-453-29 Sequence 29, Appl
18	124.5	2.7	4689	4	US-09-105-537-34 Sequence 34, Appl
19	124.5	2.7	42235	4	US-09-189-637A-1 Sequence 1, Appl
20	124	2.7	36778	4	US-09-105-537-5 Sequence 5, Appl
21	124	2.7	38506	3	US-09-320-878-19 Sequence 19, Appl
22	123	2.7	11220	4	US-09-105-537-32 Sequence 32, Appl
23	122	2.7	2151	4	US-09-627-216A-9 Sequence 9, Appl
24	122	2.7	2439	4	US-09-624-693A-18 Sequence 18, Appl
25	122	2.7	5392	2	US-08-403-852D-1 Sequence 1, Appl
26	122	2.7	5392	3	US-08-510-646B-1 Sequence 1, Appl
27	122	2.7	5392	4	US-09-231-818-1 Sequence 1, Appl
28	121.5	2.7	13842	4	US-09-105-537-30 Sequence 30, Appl
29	121.5	2.7	30001	2	US-08-125-468-1 Sequence 1, Appl
30	121.5	2.7	30001	2	US-08-474-933-1 Sequence 1, Appl
31	121	2.7	1170	3	US-09-307-621-1 Sequence 1, Appl
32	120.5	2.6	3059	4	US-09-179-558-61 Sequence 61, Appl
33	120	2.6	43280	2	US-08-804-227C-1 Sequence 1, Appl
34	119	2.6	2151	4	US-09-627-216A-7 Sequence 7, Appl
35	118.5	2.6	16836	4	US-09-147-236-1 Sequence 1, Appl
36	118.5	2.6	16836	4	US-09-147-236-10 Sequence 10, Appl
37	118.5	2.6	4403765	4	US-09-103-840A-2 Sequence 2, Appl
38	118.5	2.6	4411529	4	US-09-103-840A-1 Sequence 1, Appl
39	118	2.6	7295	2	US-08-487-826B-15 Sequence 15, Appl
40	117.5	2.6	3688	6	5246670-4 Patent No. 5246670
41	117.5	2.6	34063	4	US-09-453-702B-96 Sequence 96, Appl
42	117.5	2.6	4411529	4	US-09-103-840A-1 Sequence 1, Appl
43	117	2.6	8982	3	US-08-976-255-5 Sequence 5, Appl
44	116.5	2.6	2623	2	US-08-973-675-1 Sequence 1, Appl
45	116.5	2.6	11219	1	US-07-642-734C-1 Sequence 1, Appl

# ALIGNMENTS

RESULT 1  
US-09-276-531-47  
Sequence 47, Application US/09276531  
Patent No. 6183968  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Yue, Henry  
APPLICANT: Reddy, Roopa  
APPLICANT: Guegler, Karl J.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276, 531  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/079, 677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:



Db 481 GAA-----GACATTGACACAGTTCTTCCAGAGAGAT 513

Qy 217 TyGlnaspGlyLysGluPheGlyIleGlyAspLeuValTyrGlyLysIleGlyPhe 236  
|||:::  
:::

Db 514 TATGAAACCTTCAACCAACTAC-----AAGGCTTC 543

Qy 237 SerTyrTrpProAlaMetValIleSerTrpLys-----AlaThrSerLysArg 252

Db 544 AGCCAGTTTGTCCAGACCCCTCATTTGTGGCCAAAMATCCAGAMATTGCTGTCCAGATG 603

Qy 253 GlnAlaMetSerGlyMetArgTyrValGlnTrpPheGlyAspGlyLysPhe----- 269

Db 604 ATGATGTTTGGTGTGCAAAATGCGGAGATTCACTAGTACCAATTAACCCCTTCAAMGCAAT 663

Qy 270 -----SerGluValSerAlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPhe 287  
|||:::  
:::

Db 664 TCTGGGGCATCAGTGGAGCTGGGGAGCAGCAGCGGTGCTGTGGTGGAG----- 714

Qy 288 AsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyrHisLeuGlu 307  
::|:::  
:::

Db 715 AGCATGTGTGACAGCCACTGAGTTGGACACACACACCTCCCTGTGTGAGAGTGCATTCCGC 774

Qy 308 LysAlaArgValArgAlaGlyLysTrpPheProSerSerProGlyAspSerLeuGlnAsp 327  
|||:::  
:::

Db 775 AAGGCCAAGACCAAGAGGGCAAAAGT-----CCCAATGCT----- 810

Qy 328 GlnLeuLysProMetLeuGlnTyrAlaHisGlyLysPheLysProThrGlyIleGluGly 347  
::|:::  
:::

Db 811 CGAGAGAAAGCC----- 822

Qy 348 LeuLysProAsnLeuThrGlnProValAlaAsnLysSerLysValArgArgAlaGlySer 367  
|||:::  
:::

Db 823 ---AAGGCGACCCCTCGTGAACCTGATGCCAGAAACCTAAACCAAGAAAGTAGTCC 879

Qy 368 ArgLysLeuGlnSerArgLysTyrGlnAsnLysThrArgArgArgThrAlaAspAspSer 387  
|||:::  
:::

Db 880 CTGAAATTCAAAGCTGGAGGTTTGGTTCACACCTGAAGATCCCTCGAGTGA----- 933

Qy 388 AlaThrSerAspTyrCysProAlaProLysArgLeuLysThrAsnCysArgLysAsnGly 407

Db 933 ----- 933

Qy 408 LysAspArgGlyAspGlnAspGlnSerArgGlnGlnMetAlaSerAspValAlaAsnAsn 427  
|||:::  
:::

Db 934 -----GATGATGACTTAAATGTGAATGTGAATCTGACTTGATGATGATCCAGTATCAT 981

Qy 428 LysSerSerLeuGlnAspGlyCysLeu-----SerCysGlyValArgLysAsnProVal 444  
|||:::  
:::

Db 982 AGCTATTCTGTTCTGTATGTTCTCACACAGCCGTATGACGCGAGCGCAAGAACTCCGA 1041

Qy 445 SerPheHisProLeuPheGlnGlyGlyLysCysGlnThrCysArgAspArgPheLeuGlu 464

Db 1042 ACCCACTAAAAAGAAAAAGCGAGAGAGAGAGTACTGTGTGATGCT----- 1092

Qy 465 LeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThrValCysCysGlnGly 484  
|||:::  
:::

Db 1093 -----TATGACACAGAC---CACAGGACTATTTGGCAGGTGTGCGACGAAAGC 1137

Qy 485 ArgGlnLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCysValGlnCysLeu 504  
|||:::  
:::

Db 1138 GGTGAGATCATCTGTGTGAT-----ACCTGTCCTCGTCTTACCAATGCTGTGCTGCTG 1191

Qy 505 GlnValLeuValIleGlyThrGlyThrAlaAlaGlnAlaLysLeuGlnGlnLupProTrpSerCys 524  
|||:::  
:::

Db 1192 GAT-----CCGACATGAGAAAGGCTCCGAGAGGAG-----TGAAGCTGC 1233

Qy 525 TyrMetCysLeuProGln-ArgCysHisGlyValLeuArgArg-----Ar 539  
|||:::  
:::

Db 1234 CCACACTCCGAGAAAGGAGATTCAGTGGGAAGCTTAAGAGACATTCGAGAGGTGAG 1293

Qy 539 GlyAspTrp 542  
|||:::  
:::

```

DB 1294 GAGATCCTGG 1303

RESULT 3
US-09-249-181A-1
: Sequence 1, Application US/09249181A
: Patent No. 6440679
: GENERAL INFORMATION
: APPLICANT: Seelig, Hans Peter
: APPLICANT: Renz, Manfred
: TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
: FILE REFERENCE: 8484-0059-999
: CURRENT APPLICATION NUMBER: US/09/249,181A
: CURRENT FILING DATE: 1999-02-12
: PRIOR APPLICATION NUMBER: US 08/913,832
: PRIOR FILING DATE: 1998-01-12
: PRIOR APPLICATION NUMBER: PCT/DE96/00444
: PRIOR FILING DATE: 1996-03-08
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 6328
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)...(5736)
: US-09-249-181A-1

Alignment Scores:
Pred. No.: 0.00229 Length: 6328
Score: 148.50 Matches: 108
Percent Similarity: 34.57% Conservative: 87
Best Local Similarity: 19.15% Mismatches: 193
Query Match: 3.25% Indels: 176
DB: 4 Gaps: 26

US-09-720-086-8 (1-853) x US-09-249-181A-1 (1-6328)

Qy 14 AAlagiGlyArGluAspSerIle-----LeuValAsnGlyAlaCyS----- 27
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37 GCGGCACTAGAGAGAGAGATGATGATGACACTTTTGAACAACAGCTCGCCACCAC 96
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 28 SerAspGlnSerSerAspSerProProlIleGluAlaIleArgThrProGlnIleArg 47
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 97 CCAGAAATATGAAGAGACCAGAAAGAGATTTCAGAAAACAGACATCCAAAGCTCAAG 156
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 48 GlyArGlySerSerSerArg-----LeuSerIAsrGluVal 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 157 AAGAAAGAAAAGAACTAGAAACCTCGGAGACCTTAAATCCCTTAAGACCAAGCCCAAAA 216
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 61 SerSerLeuLeuSerTyThrGlnAspLeuThrGlyAspGlyAspGlyIuAspGlyAsp 80
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 217 AAGAGACGTATGCTCTTATGCGGACGCTG--GGGACAGCTCGGGAGAGGG-- 267
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 81 GlySerAspThrProValMetProIysLeuPheArgGluThrArgThrArgSerGluSer 100
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 268 -----CCAGAGTTTGTGAGAGAGAGAGAAAGAGGAGCTCGCCTCAGACAGT 315
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 101 ProAlaValArgThrArgAsnAsnAsnSerValSerSerArgIuArgIAsnArgProSer 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 316 GAGGCG-----AGCGACTATATCTCTCGGCAAGAAAGAAAGAAAGAAAGAAAGAAAG 363
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 121 ProArgSerThrArgGlyArgGlnGlyArgGlnIAsnIAsnIAsnIAsnIAsnIAsnIAsn 140
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 364 CCTAAGAAAGAGAAAGAAAGCAATCCAGCGCAAGAGAGAGAGAGAGAG----- 411
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 141 ProAlaThrArgSerLeuArgArgArgAlaThrAlaSerAlaGlyThrProTyrProSer 160
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 411 ----- 411
Qy 161 ProProSerSerTyLeuThrIleAspLeuThrAspAspThrArgIAsnThrIAsnGlyThr 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 412 -----GAGATATATGATGATGATGATGATTCAAAGAG 441

```

QY 181 ProGlnSerSerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMet 200  
Db 442 CTTAAATCATCT-----GCTCAGCTCTCGGAAGAC-----TGGGGCATG 480  
QY 201 GluSerProGlnValGluAlaAspSerGlyAspGlyAsp-----SerSerGlu 216  
Db 481 GAA-----GACATTGACCACGCTGTTCTCAGAGGAGAT 513  
QY 217 TyrGlnAspGlyGlyGluPheGlyGlyLeuValTrpGlyGlyIleGlyPhe 236  
Db 514 TATCGAACCTCACCACCTAC-----AAGCCCTTC 543  
QY 237 SerTrpTrpProAlaMetValValSerTrpLys-----AlaThrSerLysArg 252  
Db 544 AGCAGTTTGTTCAGACCTCATTTGCTGCCAAAATCCCAAGATTGCTGTCTCCAAGATG 603  
QY 253 GlnAlaMetSerGlyMetArgTrpValGlnTrpPheGlyAspGlyLysPhe----- 269  
Db 604 ATGATGGTTTGGGTGCAAAATGCGGGAGTTTCAGTACCAATAACCCCTTCAAGGCGAGT 663  
QY 270 -----SerGluValSerAlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPhe 287  
Db 664 TCTGGGGCATCAGTGGCAGCTGCGCGACGACGAGCGGTAGCTGTGGTGGAG----- 714  
QY 288 AsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyrHisAlaLeuGlu 307  
Db 715 AGCATGGTGACAGCCACTGAGTTGACACCAACCTCCCTGTGGAGTGCCTATCCGC 774  
QY 308 LysAlaArgValArgAlaGlyLysThrPheProSerSerProGlyAspSerLeuGluAsp 327  
Db 775 AAGGCCACGACCAAGGCGGCAAGGT---CCCAATGCT----- 810  
QY 328 GlnLeuLysProMetLeuGluTrpAlaHisGlyGlyPheLysProThrGlyLeuGly 347  
Db 811 CGAGGAGGCC----- 822  
QY 348 LeuLysProAsnAsnThrGlnProValValAsnLysSerLysValArgAlaGlySer 367  
Db 823 ---AAGGGCAGCCCTCGTGTACCTGATGCCAAGAGCCCTAAACCCCAAGAAAGTAGTCTCC 879  
QY 368 ArgLysLeuGluSerArgLysTyrGluAsnLysThrArgTrpAlaAspAspSer 387  
Db 880 CTGAAATCAAGCTGGGAGGTTTGGTTCCAGCGTAAGAGATCCTCGAGTGA----- 933  
QY 388 AlaThrSerAspTyrCysProAlaProLysArgLeuLysThrAsnCysTyrAsnAsnGly 407  
Db 933 ----- 933  
QY 408 LysAspArgGlyAspGluAspGlnSerArgGluGlnMetAlaSerAspValAlaAsnAsn 427  
Db 934 -----GATGATGACTTAGATGTGGAATCTGACTTCGATGATGCCAGTATCAAT 981  
QY 428 LysSerSerLeuGluAspGlyCysLeu-----SerCysGlyValGlyAsnProVal 444  
Db 982 AGCTATTCTGTTCTGATGTTCTCCAGCCGCTAGTAGCCGCGCGCAAGAACTCCGA 1041  
QY 445 SerPheHisProLeuPheGluGlyGlyLeuCysGlnThrCysArgAspArgPheLeuGlu 464  
Db 1042 ACCACTAAAGAAAGAAAGGAGGAGGAGGAGGAGGAGTGTGCTGGATGGT----- 1092  
QY 465 LeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThrValCysCysGluGly 484  
Db 1093 -----TATGAGACAGAC---CACACGAGACTATGCGAGGTTGCGCAGCAAGGC 1137  
QY 485 ArgGluLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCysValGluCysLeu 504  
Db 1138 GGTGAGATCATCTGTGTGAT-----ACCTGTCCCGCTGTACCATCATGCTGCTGCTG 1191  
QY 505 GluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnGluProTrpSerCys 524  
Db 1192 GAT-----CCGACATGGAAGGCTCCCGAGGGCAAG-----TGGAGTGC 1233

QY 525 TyrMetCysLeuProGln-ArgCysHisGlyValLeuArgArg-----Ar 539  
Db 1234 CCACACTCGGAGGAAGGAGCATCCAGTCGGGAAGCTTAAGAGGACAATTCGGAGGCTGAG 1293  
QY 539 gLysAspTrp 542  
Db 1294 GAGATCTCTGG 1303  
RESULT 4  
US-08-913-159-12  
; Sequence 12, Application US/08913159  
; Patent No. 6300109  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Plasmid-derived type II  
; TITLE OF INVENTION: restriction-modification systems from *Lactococcus lactis*  
; NUMBER OF SEQUENCES: 14  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913,159  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 0179/95  
; FILING DATE: 17-FEB-1995  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2355 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: *Lactococcus lactis* subsp. cremoris  
; STRAIN: W39  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 744..1283  
; IDENTIFICATION METHOD: experimental  
; OTHER INFORMATION: /codon\_start= 744  
; OTHER INFORMATION: /product= "Lladii restriction endonuclease"  
; OTHER INFORMATION: /evidence= EXPERIMENTAL  
; OTHER INFORMATION: /gene= "ORF"  
; OTHER INFORMATION: /number= 1  
; OTHER INFORMATION: /standard\_name= "Gene coding for R.Lladii"  
; OTHER INFORMATION: /label= "r-lladii"  
; OTHER INFORMATION: /note= "The first ten amino acids in this sequence may be  
; OTHER INFORMATION: doubtful. However, from base 773 this reading frame gives a h  
; OTHER INFORMATION: homology with the Bsp6I endonuclease"  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1392..2342  
; IDENTIFICATION METHOD: experimental  
; OTHER INFORMATION: /codon\_start= 1392  
; OTHER INFORMATION: /product= "Lladii methylase"  
; OTHER INFORMATION: /evidence= EXPERIMENTAL  
; OTHER INFORMATION: /gene= "ORF"  
; OTHER INFORMATION: /number= 2  
; OTHER INFORMATION: /standard\_name= "Gene coding for M.Lladii"  
; OTHER INFORMATION: /label= "m-lladii"  
; OTHER INFORMATION: /note= "The sequence shows 60 % identity and 76 % similarity  
; OTHER INFORMATION: with the Bsp6I methylase."  
US-08-913-159-12  
Alignment Scores:  
Pred. No.: 0.000753 Length: 2355  
Score: 146.00 Matches: 74  
Percent Similarity: 36.39% Conservative: 49



Query Match:	21.89%	3.20%	Mismatches:	125
DB:	4	Gaps:	Indels:	90
				17
US-09-720-086-8 (1-853) x US-08-913-159-12 (1-2355)				
Qy 575	IIleArGVAlleuSerleuPheAspGIYIIeAlthnGlyTYrLeuValleuLysGluLeu			594
Db 1395	TTGAAATTCGCTTCCTTTTTCCTGGCCGAGTTGGCGGAATGATTGCTTTGAAATGCA			1454
Qy 595	GIYIIleLysValGIYsTYrValAlAserGIuValCYsGluLyserIIeAlaValGIY			614
Db 1455	GGTTCAAAACA-----AAATATGCAATGAATTT-----GAAATATATGCTGCGAT			1502
Qy 615	ThYrValIshIsGIuGIYAsnIIeLYsTYrValAsnAspValArgAsnIIeThrLYsLYs			634
Db 1503	ACTTTTGAATGAACCTTGACGTTAAGGTA-----GACCGACGTAATATGAATGATGA			1556
Qy 635	AsnIIeGIuLurpGIYrProPheAspLeuValIIeGIYLYserProCYAsnAspLeu			654
Db 1557	CAACCTGATGAATATCCAGATTTTGATATATGATATGACGATTTTCCTTGCCAAACCTTT			1616
Qy 655	SeArAsnValAsnProAlaArgLYsGIYLeu-----TYrGIuGIYThrGIYArgpHe			672
Db 1617	TCT---ATTCGTGGTTATCGTCAAGCTTTAAGATGAACAAGTCGAGTATCTTTT			1673
Qy 673	PheGIuPheTYrHisLeuLeuAsnTYrSerArgProLYsGIuGIYAspAspArgProHe			692
Db 1674	TTTGACCTGTTGCTATTTTGAACAACAAAACCTCGTTGCA-----			1718
Qy 693	PheTrpMetPheGIuAsnValAlaIleMeLYsValGIYAspLYsArgAspIIeserArg			712
Db 1719	-----TTCTTGAAATATGTTAAAAATCTTGTTCTCAAGATAGCGGGAACATTGGA			1772
Qy 713	PheLeuGIuCYAsnProVal-----MeIIesAspAlaIIeLYs			725
Db 1773	GTTAT---TGTTCGAGTTTGAAGAAAGACTRAGGTACAGATCTTTTTCAGTGTAT			1829
Qy 726	ValSerAlaAlaHisArgAlaArgTYrPheTrpGIYAsnLeuProGIYMeAsnArgPro			745
Db 1830	GCTTCTGAA-----TATGAAATATATCCT---CAAAATAGAGAA			1865
Qy 746	-----ValIIeAlAserLYsAsnAsp-----			752
Db 1866	CGTATCTATATTTGTTGCTTTCAAAAATATAAAAAGATTATGCAATTTTGAACTACAAA			1925
Qy 753	-----LYsLeuGIuLeuGIuAspCYsLeuGIuTYrAsnArg-----			764
Db 1926	TCTATACCTTTAAAAACAAGATTCACGATTTATGATTTTTCCTAAAAACAAGACAT			1985
Qy 764	-----			764
Db 1986	AAGTTCTACTACTCTCGTGAAGAAATATTTTGTGATGAGCTTAAAAAATAATGACT			2045
Qy 765	-----IIeAlaLYsLeuLYsLYsValGIuThrIIeThrThLYsSerAsnSer			780
Db 2046	AATCAAGCACTACATCATCATCGTGCGTAGAGTTTATGTATAGAGAAAAACAAGTATTTA			2105
Qy 781	IIleLYsGIu-----GIYLYsAsnGIuLeuPheProValAl			792
Db 2106	GTACCAACACTAACGCGCTAATATAGGAACAAGTGGGCATATAT-----GTGCTTATATTC			2159
Qy 793	MeArAsnGIYsGIuAspValleuTrpCYeThrGIuLeuGIu-----ArgIIePheGIY			810
Db 2160	CTTACATATACGGAGATATTCGTAATTTAAACCAAGAGAAAGCTTTTAAGCTTCAAGT			2219
Qy 811	PheProValHisTYrThrAspValSerAsnMeGIYArgGIYAlaArgIIuLYsLeuLeu			830
Db 2220	TTCCCAAAAAGATAT---AAACTTCCAAAACCAAGTAAATAGCGGATTTATATAAACAAG			2276
Qy 831	GIYAsnArgTrpSerValProValIIeArgHisLeuPheAlaProLeuLYsAsp			848
Db 2277	GGAAACAGGTGTTGATGATCCAGTTATGAAAGAATGCAAAAAAATCTTGCAAGT			2330

```

RESULT 5
US-09-484-970B-8
; Sequence 8, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmuth, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENDS
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 8
; LENGTH: 3522
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6426186 249096.3
; NAME/KEY: unsure
; LOCATION: 556, 658, 663
; OTHER INFORMATION: a, t, c, g, or other
; US-09-484-970B-8

```

Alignment Scores:

Pred. No.:	0.00206	Length:	13
Score:	144.50	Matches:	35
Percent Similarity:	34.78%	Conservative:	81
Best Local Similarity:	21.74%	Mismatches:	23
Query Match:	3.16%	Indels:	17
DB:	4	Gaps:	32

US-09-720-086-8 (1-853) x US-09-484-970B-8 (1-3522)

Oy	28	SerAspIleSerSerAspSerProProlIeuGluAlaIleArgThrProGluLeuArg	47
Db	766	ACAGAACCAATTCACCCCAGAACCCTTTCACACTAAGATTCCACGAACAATAAGTA	822
Oy	48	GlyArgArgSerSerSerArgLeuSerLysArgGluValSerSerLeuIleuSerYrThr	67
Db	823	-----GCAAAGACACTCAGGGGCCACACAGATTTTATACTGTGTGAGGCCAGAAAC	876
Oy	68	GlnAspLeuThrGlyAspGlyAspGlyGluAspGlyAspGlySerAspThrProValMet	87
Db	877	-----TCTGACAGCGCACACATC	894
Oy	88	ProLysLeuPheArgGluThrArgThrArgSerGluSerProAlaValArg---ThyrArg	106
Db	895	AGACCTTGTTGGAATAGCAACMACTAACAAACCTTACTAGGCCCAAACCCAGTAGGATGCC	954
Oy	107	AsnAsnAsnSerValSerSerArgGluArgHisArgProSerProArgSerThrArgGly	126
Db	955	AGTGGGAATGGAGTGGCAACAGGGGTCACAGCAAGCACCCAGGCCA-----TCAGGT	1005
Oy	127	ArgGlnGlyArgAsnHisValAspGluSer---ProValGluPheProAlaThrArg---	144
Db	1006	GCTGATAGAAATGTATCATGGACCTCTACCCACCACCTAAAAAACCGAGGACTGCGGC	1065
Oy	145	--SerLeuArgArgArgAlaThrAlaSerAlaGlyThrProTrpPro-----	159
Db	1066	CCACCTTTGCCACCCACAGACTTACACACCACGAGAAAACTTTTACACCAAAATATATGTC	1125
Oy	160	----SerProProSerSerYrIleuThrIleAspLeuThrAspAsp-ThrGluAspThr	177
Db	1126	ACTGAAAGCCAGGAAGTGCAGGAATTCATTATCATCGGGCCCAATTAATCATCACACCCCT	1185
Oy	177	rHisGlyThrProGluSerSerSerThrProTyrlaArgLeuAlaGluAspSerGlnGly	197
Db	1186	GAGGTCAACACCCAGGCGCTTACTGAACTCCCTTTGGAGAGAAATAGAGACAGATATTAAGCA	1245

QY 197 nGlyMetGluSerProGlnValGluAlaAspSerGlyAsp-----G1 212  
Db 1246 A-----CCAAAGTTCCTGCC--TCTGGAGAGAACTGGAAAATAAAC 1287  
QY 212 yAspSerSer-GlutYrGlnAspGlyLysGluPheGlyileGly-----A 227  
Db 1288 TGACTTTAGCTCAAGCCCAACAAGAGAACTGATCTCTTGGGAAGCCCAAGATTCAAAGG 1347  
QY 227 sPLeuValTrpGlyLysileGlyPheSerTrpProAlaMetValValSerTrp- 246  
Db 1348 ACCTCATGTGCATACATCCAAAAGCCTGACACAGCTCCTGCTCCATCTACTGACTGT 1407  
QY 247 --LysAlaThrSerLysArgGlnAlaMetSerGlyMet----- 258  
Db 1408 CAACGGTTTCCCAAGAGAGGCCACAGAGGGAATGCCACGCCACCACAGAAACC 1467  
QY 259 -----ArgTrpValGlnTrpPheGlyAspGlyLysPheSerGluValSerA 274  
Db 1468 ACCCAACCACTCACTGGTGCACCGTGAAGGGTCCCTTCATTTGTCTATCTGGACT 1527  
QY 274 laAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsnL 294  
Db 1528 GGGAAAGCCACTAAAT-----GACACTGTCTACTGAATATG 1563  
QY 294 yLeuValSerTrpArgLysAlaMetTrpHisAlaLeuGluLysAlaArgValArgAlaG 314  
Db 1564 AAGTTATATCCAGAGAAATGGGTCACTTTCAGTGGGAAGAACCAAGTCCATTCAAATGACAA 1623  
QY 314 lYlYsThrPheProSer-----SerProGlyAspSerLeuGluAspGlnLeuL 330  
Db 1624 ATCAGACATTTTCCACAGTAGAAAATCTGGAACCAACACAGAGTTATGAATTCAGGTGA 1683  
QY 330 yProMetLeuGluTrpAlaHisGlyLysPheProThr-----GlyileGluG 347  
Db 1684 AACCCCAAAACCCGTTGGTGAAGCCCGGTGACGACACAGTGCATTCAGTACTGAAT 1743  
QY 347 lYleuLysProAsnAsnThrGlnProVal----- 356  
Db 1744 CAGCGGACCCAAAGAGTAGTGAGCCAGTTCTTCAGGAAGAGATGCCATCTGGACTGAAA 1803  
QY 357 -----ValAsnLysSerLysValArgArgAlaG 366  
Db 1804 GACCCCTTTAATTCAGACTCTTTACTCAGAGTGTAAAGGGCAACAATATGTCAAAGGACAT 1863  
QY 366 lYserArgLysLeuGluSerArgLysTrpGluAsnLysThrArgArgArgThr---AlaA 385  
Db 1864 GGTATAAAAAATTTGTAGGAGTGCAGCTGTGCAACTCTCTCAGATACAAAGATTACTTGA 1923  
QY 385 sPAspSerAlaThrSerAspTrpCysProAlaProLysArgLeuLysThrAsnCysTrpA 405  
Db 1924 GCGACTCCTCCACAGGAAATTT-----TATA 1950  
QY 405 sPAsnGlyLysAspArgGlyAsp-----G 413  
Db 1951 ACATAGGTGATCAGAGGGGCCATGGAGAGATCACTGCCAGTTTGTGGATTCATTTTAG 2010  
QY 413 luAspGlnSerArgGluGlnMetAlaSerAspValAlaAsnAsnLysSerSerLeuGluA 433  
Db 2011 ATGGACGCACCTGGGCGACCACTCACTTCTGAC-----CAGTTACCACTCAAAG 2058  
QY 433 sPGLYCysLeuSerCysGlyArgLysAsnProValSerPheHisProLeuPheGluGlyG 453  
Db 2059 AAGGTATTTCAGAGCACTTCGCCAGCAACCTGTCCAATTTGGAGAAATA-----GGTG 2112  
QY 453 lYleuCysGlnThrCysArgAspArgPheLeuGluLeuPheTrpMetTrpAspAspG 473  
Db 2113 GTCACACCCAAATC-----AAATTATGTTTCAGTGGTATGA-ATGTGGGAC----- 2155  
QY 473 lYTYrGlnSerTrpCysThrValCysCysGluGlyArgGluLeuLeuLeu-----CysA 491  
Db 2156 --TACAATTCCTGGAAAATGTAGATGCTGCACAAAGTTTACCTTCTGTTTCATCATTTGCA 2213  
QY 491 erAsnThrSer-----CysCysArg-----CysPheCysValGluCysL 504

Db 2214 AACAAAATCATTTGAAATACTATGCCCATTCATTTAAAGCTATTTTGTACTATGTA 2273  
QY 504 euGluVal-----LeuValGlyThrGlyThr-----AlaAlaG 515  
Db 2274 TAAAGTCTACAATCTAAATTAATGCAATCACTAGATGTTTATTATAGAAAGATTGCTG 2333  
QY 515 luAla-----LysLeuGlnGluPro----- 521  
Db 2334 AGAGTATTATTCAGGTTTTCACAAAGTCCATTTTAAAGAACCAAGATCTGATGTTAACAG 2393  
QY 522 -----TrpSerCysTrpMetCysLeuProGlnArgCysHisGlyValLeuArg 537  
Db 2394 AATAACATTTTGGGAGAGCTGGCTCCCTATT-----CATGGTATTTTAAAG 2440  
RESULT 6  
US-09-041-886-18  
; Sequence 18, Application US/09041886  
; Patent No. 6235872  
; GENERAL INFORMATION:  
; APPLICANT: Bredesen, Dale E.  
; APPLICANT: Rabizadeh, Sharroz  
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence  
; TITLE OF INVENTION: Polypeptides and Methods of Use  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/041.886  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 2626  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4481 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 163..4099  
; US-09-041-886-18  
Alignment Scores:  
Pred. No.: 0.0144 Length: 4481  
Score: 137.50 Matches: 108  
Percent Similarity: 33.46% Conservative: 63  
Best Local Similarity: 21.14% Mismatches: 211  
Query Match: 3.01% Indels: 132  
DB: Gaps: 18  
US-09-720-086-8 (1-853) x US-09-041-886-18 (1-4481)  
QY 11 GluGluAspAlaGlyGlyArgGluAspSerileLeuValAlaGlyAlaCysSerAspGln 30  
Db 1375 GAAGAAAATTATGTT-----GTAGTGCTACGTATGATGACGATTTA 1416

Qy 31 SerSerAspSerProProlLeuGluAlaIleArgThrProGluIleArgGlyArgArg 50  
 Db 1417 TCTTCGATATACAGTCCCTTTAGAAAAGAGATACAGAAAGATTTTAAACGGGAA 1473  
 Qy 51 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuSerTyrThrGlnAspLeu 70  
 Db 1474 GCAAGGCGAAACAGTATAGCAGAAAGATTTAGTCCAGTCCAGTACAAAGCTGAGT 1533  
 Qy 71 ThrGlyAspGlyAspGlyGluAspGlyAspGlySerAspThrProValMetProLysLeu 90  
 Db 1534 GCCCTGGAATAATGATGATAGAGAGAGAGAAATAATACAGCAAGTT----- 1581  
 Qy 91 PheArgGluThrArgThrArgSerGlySerProAlaValArgThrArgAsnAsnSer 110  
 Db 1582 ---CAGAGAAATTCAG 1638  
 Qy 111 ValSerSerArgGluArgGlyArgProSerProArgSerThrArgGlyArgGlyArg 130  
 Db 1639 ATTCTCTCGAGCAAAAGAAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1698  
 Qy 131 AsnHisValAspGlySerProValGluPheProAlaThrArgSerLeuArgArgAla 150  
 Db 1699 CCGCGATGGCCAG-----CCTGAGTCGGCTCCATGCCATCAAGATCC 1743  
 Qy 151 ThrAla-----SerAla 154  
 Db 1744 ACTTCTCACTGAGATTGACCCGAAATCTGTGTCAGACCAAGAGATGATTAATGGA 1803  
 Qy 155 GlyThrProThrProSer-----ProProSerSerTyrLeu 166  
 Db 1804 GGTGTCTCTGCGCATCGCTTGGCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1860  
 Qy 167 ThrIleAspLeuThrAspAspThrGlnAspThrHisGlyThrProGlnSerSerThr 186  
 Db 1861 -----CACTCAGCTCCCACTCTTCCAC-TCGAGCGCCAC 1898  
 Qy 187 ProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMetGlySerProGlnValGlu 206  
 Db 1899 CCGTACAGCGCGCGCGCTCCAG---CGCCCTCGCGCGCGCGCGCGCGCGCTCTCACC 1955  
 Qy 207 AlaAspSerGlyAspGlyAspSerSerGluTyr-----Gln 218  
 Db 1956 CTCTGCTCATGCTCTCCAGCTCTGCTCTACTGCTTAAAGCAGATGCTTCAGAAAG 2015  
 Qy 219 AspGlyLysGluPheGlyIleGlyAspLeuValTyrGlyLysIleGlyPheSerTyr 238  
 Db 2016 GCCTTCAGAGATGTCCTCCAAAGCGCCAGCAGATCTCGAAATACAGAGTTTCTGCTG 2075  
 Qy 239 -----TyrProAlaMetValVal-----Ser 245  
 Db 2076 GAGGGCTTCATTCAGTGGCTTGAATTTGTATCCCAACCAACCCAGTAAAGCAGC 2135  
 Qy 246 TrpLysAlaThrSerLysArgGlnAlaMetSerGlyMetArgTyrValGlnTyr----- 263  
 Db 2136 TACCTCTCCAGTAAAGAGACAGATCTCTCGGGGGA-CTGGTCAATCAAGTGTCACTG 2194  
 Qy 264 ---PheGlyAsp----- 266  
 Db 2195 GGGTTCAGATTTATCCCTTAAACTCATAGACCAGGTCTCCAGACAGACAGATATTG 2254  
 Qy 267 -----GlyLysPheSerGluValSerAlaAspLysLeuValAlaLeuGly 281  
 Db 2255 GAAATACCCCAAGGCGAGGCTTCTCTCTCCCAAGCTGATTTATTCCAACTGGAAG 2314  
 Qy 282 LeuPheSerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAla 301  
 Db 2315 CTGTTGCAATGCTATTTCCAGTGCATCTCAAGCCCTGTAAGTCTTCATCGAAGACAG 2374  
 Qy 302 MetTyrHisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSerPro 321  
 Db 2375 CTGTATCCCTTTAGTGAAGGCTAAAGAT-----CCA 2407

Qy 322 GlyAspSerLeuGluAspGlnLeuLysProMetLeuGluThrAlaHisGlyPheLys 341  
 Db 2408 GGGTTCAGATTCAGAG-CAGAACTCTCT-----GCAGGGAATATA 2448  
 Qy 342 ProThrGlyIleGlyLysLeuLysProAsnAsnThrGlnProValValAsnLysSerLys 361  
 Db 2449 -----GAAATATTAAACCAATGAACATACACCTAGCTTCTCAAAAGCTGAA 2496  
 Qy 362 ValArgArgAlaGlySerArgLysLeuGlySerArg-----Lys 374  
 Db 2497 AACAAAGTATATACACAGTGTCTTTCGAACATGAAACAGATTGATTTAAAGAA 2556  
 Qy 375 TyrGluAsnLysThrArgArgThrAlaAspSerAlaThrSerAspTyrCysPro 394  
 Db 2557 TTTAAGATGATTTTAAAGTTTACAGCAAGTCTTCTGAACTATGATCAACTACTA 2616  
 Qy 395 AlaProLysArg-----LeuLysThrAsnCysTyrAsnAsn 406  
 Db 2617 AACAAATATAG 2676  
 Qy 407 GlyLysAspArgGlyAspGluAspGlnSerArgGluGlnMetAlaSerAspValAlaAsn 426  
 Db 2677 GCTAAGATTTCTTCAATTGAAATATGACAGCACTGTACCACTGGCAGCAGACGCC 2736  
 Qy 427 AsnLysSerSerLeuGluAspGlyCysLeuSer 437  
 Db 2737 AATAGCCCGACATTTCCCTTCAATTAAGT 2769

RESULT 7  
 US-09-105-537-5/c  
 ; Sequence 5, Application US/09105537A  
 ; Patent No. 6265202  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D.H.  
 ; APPLICANT: Liu, H.  
 ; APPLICANT: Xue, Y.  
 ; APPLICANT: Zhao, L.  
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
 ; FILE REFERENCE: 600 438US1  
 ; CURRENT APPLICATION NUMBER: US/09/105,537A  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 5  
 ; LENGTH: 36778  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces venezuelae  
 ; US-09-105-537-5

Alignment Scores:  
 Pred. No.: 0.493 Length: 36778  
 Score: 137.50 Matches: 111  
 Percent Similarity: 32.45% Conservative: 49  
 Best Local Similarity: 22.52% Mismatches: 170  
 Query Match: 3.01% Indels: 166  
 DB: 4 Gaps: 21

US-09-720-086-8 (1-853) x US-09-105-537-5 (1-36778)  
 Qy 10 GlyLysGluAspAlaGlyArgGluAspSerLeuValAsnGlyAlaCysSerAsp 29  
 Db 31747 GGGGTGTGGGGGGGGGTGGTGGCGGGGATCC-----CGGGGTGGCGGGCGGTGG 31697  
 Qy 30 GlnSerSerAspSerProProlLeuGluAlaIleArgThrProGluIleArgGlyArg 49  
 Db 31696 GCAGAGCCGCGGTGGTCAAGTGTACGGGGCGCGAGAGCC-----ATCCGATCAGG 31643  
 Qy 50 ArgSerSerSerArgLeuSerLysArgGluValSerSerLeuLeuSerTyrThrGlnAsp 69  
 Db 31642 GCCTCGGGGCTCAGG-----TCGTGATC-----GAC 31616  
 Qy 70 LeuThrGlyAspGlyAspGly-----GluAspGlyAsp 80

Db 31615 GCCTCGGCTCCGACACAGGCTCGCGCGCCGCTCCGAAACCCCGGCTCG 31556  
Qy 81 GlySerAspThrProValMetProLysLeuPheArgGluThrArgThrArgSerGluSer 100  
Db 31555 GGCTCG-----ATCCCGTG-----AGCGCGAGGACGGTGTGAGGACC 31517  
Qy 101 ProAlaValArgThrArgAsnAsnAsnSerValSerSerArgGluArgHisArgProSer 120  
Db 31516 CCCTCGTCCCGCAGCGGTGAGGGGCGAGTTCCGCGCAGGGCCGCGGCGACCCGCTCC 31457  
Qy 121 ProArgSerThrArgGlyArgGlnGlyArgAsnHisValAspGluSerProValGluPhe 140  
Db 31456 CAGTCGTCGGGGCGGCTCGCGCGTCCGAGGTACGCTCG----- 31415  
Qy 141 ProAlaThrArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProTyrProSer 160  
Db 31414 -----TGGAGGTGCGCGCGAGTGCAGGCGGCGTGGGTGCTGGAAG 31373  
Qy 161 Pro-----ProSerSerTyrLeuThrIleAspLeuThrAspThrGluAspThrHis 178  
Db 31372 ACCGTGGTGGCGGAGCGCTGTCGCCACCCAGCGTGGAGAGCTG-GTTGCGGAGCTGCAC 31314  
Qy 179 GlyThrProGlnSerSerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGly 198  
Db 31313 GGC----- 31311  
Qy 199 GlyMetGluSerProGlnValGluAlaAspSerGlyAspGlyAspSerSerGluTyr-Gl 218  
Db 31310 -----CGTCAGGAGTGCAGAGCGAGCTCGTGAAGGACACGCGCGGGGCCACCCG 31260  
Qy 218 nAspGlyLysGluPheGlyIleGlyAspLeuValTrpGlyLysIleLysGlyPheSerTr 238  
Db 31259 GTCCGGGGAGGAA----- 31247  
Qy 238 pTrpProAlaMetValSerTrpLysAlaThrSerLysArgGlnAlaMetSerGlyMe 258  
Db 31246 -TGCCCGAGTACGCGCGCGCGGTGGTACGAGACAGGAGTACAGGCGCC-----GGCCG 31194  
Qy 258 tArgTrpValGlnTrpPheGlyAspGlyLysPheSerGluVal-----SerAlaAs 275  
Db 31193 GCGC-----TCGGGCTCGGGGAGCGCGGTGTCGCGCGCACGCGCGA 31152  
Qy 275 p----- 275  
Db 31151 CGACTGCCGCTCGCGGCCACGCGCGCTCGCGGAGCGCGGCGACCGAGCGGTGCGGC 31092  
Qy 276 -----LysLeuValAl 279  
Db 31091 GAGCGCTCGCGGCTCCGGGACGCGCTCGACAGAGGCTGGGACGGACACCGTGA 31032  
Qy 279 aLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrAr 299  
Db 31031 CGGGGCGGGAACCGCTCCCGTCCGACATCGGCCACGGGACGAGAGTCTCGTGTGGCT 30972  
Qy 299 gLysAlaMetTyrHisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSe 319  
Db 30971 CAGGGGCTTG-----GCCAGTTCGTCAGGGCGCGGTGCGGG----- 30935  
Qy 319 rSerProGlyAspSerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGl 339  
Db 30934 -----CTCATCGACCGATGCGCGACGCTGCCAGTACGCGTCTCGCGC 30891  
Qy 339 yPheLysProThrGlyIleGluGlyLeuLysProAsnAnThrGlnProValValAsnLys 359  
Db 30890 GCGCGGCGCATGCGCTCGCGGCCACAGAGGCCCGACGAGTCCGCGCTCTCGCC 30831  
Qy 359 sSerLysValArgAlaGlySerArgLysLeuGluSerArgLysTyrGluAsnLysTh 379  
Db 30830 CCGGGCGCGCGCGCGCGCGCGCGCTCG----- 30800  
Qy 379 rArgArgThrAlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLe 399  
Db 30799 -----AGTGGGCGTGGCGCGCGGTAGAGCGCTCGTGGCTCGCGCTGCC----- 30755

Qy 399 ulysThrAsnCysTyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluGl 419  
Db 30754 -CAGACCCCGCGTTCGAGGAGTA-GAGGACAAAGGCTCCAGCGGAGTCCCGCGGAGCA 30697  
Qy 419 nMetAlaSerAspValAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerCys-- 438  
Db 30696 GGTGTCGAGGACCTCG-----CGCGCTCTCTTCGCGGCCAGGATCGGGCGATGTCC 30641  
Qy 439 -----GlyArgLysAsnProValSerPheHisProLeu 449  
Db 30640 TCCGGCGCGGTGAGCTCCAGCGATCGCGCGCGGTG 30604  
RESULT 8  
US-09-320-878-19/c  
; Sequence 19, Application US/09320878A  
; Patent No. 6117659  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary C.  
; APPLICANT: McDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 300622002120  
; CURRENT APPLICATION NUMBER: US/09/320,878A  
; EARLIER FILING DATE: 1999-05-27  
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908  
; EARLIER FILING DATE: 1998-08-28  
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
; EARLIER FILING DATE: 1998-05-06  
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
; EARLIER FILING DATE: 1997-04-30  
; EARLIER APPLICATION NUMBER: 60/119,139  
; EARLIER FILING DATE: 1999-02-08  
; EARLIER APPLICATION NUMBER: 60/100,880  
; EARLIER FILING DATE: 1998-09-22  
; EARLIER APPLICATION NUMBER: 60/087,080  
; EARLIER FILING DATE: 1998-05-28  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 38506  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-320-878-19  
Alignment Scores:  
Pred No.: 0.532 Length: 38506  
Score: 137.50 Matches: 111  
Percent Similarity: 32.45% Conservative: 49  
Best Local Similarity: 22.52% Mismatches: 170  
Query Match: 3.01% Indels: 166  
DB: 21 Gaps: 21  
US-09-720-086-8 (1-853) x US-09-320-878-19 (1-38506)  
Qy 10 GlyGluGluAspAlaGlyArgGluAspSerIleLeuValAsnGlyAlaCysSerAsp 29  
Db 29889 GGGGCTGTGGGGGCGGTGTCGCGGGATCG-----CGGGGTGCGCGGTGGG 29839  
Qy 30 GlnSerSerAspSerProIleLeuGluAlaIleArgThrProGluIleArgGlyArg 49  
Db 29838 CGAGGACCGCGGTGCGGTGAGTGTACGGGGCGCGAGAGCC-----ATCCGATCAGG 29785  
Qy 50 ArgSerSerArgLeuSerLysArgLysValSerSerLeuLeuSerTyrThrGlnAsp 69  
Db 29784 GCCTCGCGCTCCAGG-----TCGTCGATC-----GAC 29758  
Qy 70 LeuThrGlyAspGlyAspGly-----GluAspGlyAsp 80  
Db 29757 GCCTCGGCTCCGACCGAGGCTCGGGGCGCGCGTCCGAAACCGCGGACCCGCTCG 29698

Qy 81 GlySerAspThrProValMetProIleuPheArgIleuThrArgThrArgSerGluSer 100  
Db 29697 GGGTCG-----ATCCCGGTG-----AGCCGACGACGCTCCAGGACC 29699  
Qy 101 ProAlaValArgThrArgAsnAsnAsnSerValSerSerArgGluArgHisArgProSer 120  
Db 29658 CCCCCTGTCCTCCAGCCGATCCAGAGGCGCATTCGGCCAGGCGCCGCGACCCGCTCC 29699  
Qy 121 ProArgSerThrArgGlyArgGlnGlyArgAsnHisValaPheGluSerProValGluPhe 140  
Db 29598 CAGTCCTCGGCGCGCGCTCGCGCTCGAGTACGCTCG----- 29557  
Qy 141 ProAlaThrArgSerLeuArgArgArgAlaThrAlaSerAlaGlyThrProTrpProSer 160  
Db 29556 -----TGAGAGTGCCTCGCGCGCGCTCGCGCTCGCTGTCGCAAG 29515  
Qy 161 Pro-----ProSerSerTyrLeuThrIleAspLeuThrAspAspThrGluAspThrHis 178  
Db 29514 ACCGTGTCGCGGAGCGCTGTGCTGCGACACCGTGGAGACTCG-CTTGGCGAGCTGAC 29456  
Qy 179 GlyThrProGlnSerSerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGly 198  
Db 29455 GGC----- 29453  
Qy 199 GlyMetGluSerProGlnValGlnAlaAspSerGlyAspGlyAspSerSerGluTyr-GI 218  
Db 29452 -----CGTACGCGCGGAGCGAGCTCGGTGAAGCAGCGCGGCGGCGCACCCG 29402  
Qy 218 nAspGlyArgGlnPheGlyIleGlyAspLeuValTrpGlyLysIleLysGlyPheSerTr 238  
Db 29401 GTCCGGGGAGGAA----- 29389  
Qy 238 PTrpProAlaMetValValSerTrpLysAlaThrSerLysArgGlnAlaMetSerGlyMe 258  
Db 29388 -TGCGCGAGTACGCGCGCGCGCTGCGTACGAGCAGAGGTAGAGAGCCG-----GGCGC 29336  
Qy 258 LArgTrpValGlnTrpPheGlyAspGlyLysPheSerGluVal-----SerAlaAs 275  
Db 29335 GCGC-----TCGGGCTCGGAGGCGCGGTGATCGCGCGCAGCGCGCA 29294  
Qy 275 P----- 275  
Db 29293 CGACTGCGCGTCCGCGCAGCGCGCGCTCGCGGAGCCGCGGCGACCGAGGTGCGGC 29234  
Qy 276 -----LysLeuValAl 279  
Db 29233 GAGCGCTCGCGCGCTCCGGGACCCCTCGACGAAAGCTGGGACCGGACACCGTAA 29174  
Qy 279 AlaGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrAr 299  
Db 29173 CGCGGCGCGAAGCGCTCCAGTGCATCGCGCAGCGGCAAGAGTCTGCTGCTGCT 29114  
Qy 299 GlyAlaMetTyrHisAlaLeuGlnLysAlaArgValaArgAlaGlyLysThrPheProSe 319  
Db 29113 CAGGCGCTTCG-----GCCAGTTCGTCAGGCGCGCTCGCGG----- 29077  
Qy 319 rSerProGlyAspSerLeuGlnAspGlnLeuLysProMetLeuGlnTrpAlaHisGlyGI 339  
Db 29076 -----CTCATCGAGAGATCGCGGACGCTGCCCATCGCTCTCGCGC 29033  
Qy 339 yPheLysProThrGlyIleGlnGlyLeuLysProAsnAsnThrGlnProValaLysLys 359  
Db 29032 GCCCGGCGCATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCGCGC 28973  
Qy 359 sSerLysValaArgAlaGlySerArgLysLeuGlnSerArgLysTyrGluAsnLysTh 379  
Db 28972 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCTCG----- 28942  
Qy 379 rArgArgArgThrAlaAspAspSerAlaThrSerArgTyrCysProAlaProlLysArgLe 399  
Db 28941 -----AGTGGGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCTCGCGC----- 28897  
Qy 399 uLysThrAsnCySTyrAsnAsnGlyLysAspArgLysArgLysArgLysArgGluGI 419

Db 28896 -CACACCCCGCGCTTCAGAGAGTA-GAGGACGAAGCGCTCCAGCGGAGTCCGCGGAGCA 28839  
Qy 419 nMetAlaSerAspAlaAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerCys-- 438  
Db 28838 GGTGCTGAGAGACTCGG-----CGCGCTGCTCTTCGCGCGCGCGCGAGATGCGGCGATTC 28783  
Qy 439 ---GlyArgLysAsnProValSerPheHisProLeu 449  
Db 28782 TCCGGCGCGGTGACGTCCAGCGGATCCCGCGCGGTG 28746  
RESULT 9  
US-08-923-137-1/c  
Sequence 1, Application US/08922137  
Patent No. 6083716  
GENERAL INFORMATION:  
APPLICANT: Wilson, James M.  
APPLICANT: Fatina, Steven F.  
APPLICANT: Fisher, Krishna J.  
TITLE OF INVENTION: Chimpanzee Adenovirus Vectors  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr., P.O. Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: United States of America  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/923,137  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/024,700  
FILING DATE: 06-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: GNVEN.021CIP1USA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35524 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
US-08-923-137-1  
Alignment Scores:  
Pred. No.: 1.25 Length: 35524  
Score: 133.00 Matches: 149  
Percent Similarity: 29.31% Conservative: 60  
Best Local Similarity: 20.90% Mismatches: 222  
Query Match: 2.91% Indels: 282  
Gaps: 34  
US-09-720-086-8 (1-853) x US-08-923-137-1 (1-35524)  
Qy 102 AlaValArgThrArgAsnAsnSerValSerSerArgGluArgHisArgProSerPro 121  
Db 17568 TCAATGGCGCGTCCGAGCGCTCAACCGCAGAGCTATGCTGTATCGCCCGGATCC 17509  
Qy 122 ArgSerThrArgGlyArgGlnGlyArgAsnHisValaPheGluSerProValGluPhePro 141  
Db 17508 CGCGCTATGATGGGGATCAAGAGCAGGCGAGAAACCAACCCCGAGCTCTTGATGATGC 17449

QY 371 luSerArgLysTyrGluAsnLysThrArgArgThrAlaAspSerAlaThrSerA 391  
 Db 16416 GGGCGAAGCTCGCCACAAACTCCACAGA-----CCATCATGCCCATCTCTT 16369  
 QY 391 sp-TyrCys-----ProAla 395  
 Db 16368 CTTTCTTGCTCTTTTGTGACCCGCTTGATTTTGGCGGCTTTTTTTCATCTCCTCACCAGCG 16309  
 QY 396 ProLysArgLeuLysThrAsn-----Cys-----403  
 Db 16308 GACCGTAGATTTCCAGCGCGATGACCTCGGAGCATCTCTTCTCTGTATTTGGCTTGGAC 16249  
 QY 404-----TyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluGln 419  
 Db 16248 TCCTCGCTGGGACACAAATCGGA-----16225  
 QY 420 MetAlaSerAspValAlaAsnAsnLysSerSerLeuGluAspGlyCys-----LeuSer 437  
 Db 16224-----GACTGCTCAGTATCTTCTAAGAGCGAGGGAGCGGTGCGCAGCGGCACCG 16174  
 QY 438 CysGlyArg-----LysAsnProValSerPheHisProLeuPheGluGlyGly-----453  
 Db 16173 GCTGGCGGTGGTGGCGTCCGCCACCGATACATTCCTCTTCGGTTGGCCATGT 16114  
 QY 454-----LeuCysGlnThrCysArgAspArgPhe 462  
 Db 16113 TGGCAATGGCGCGTGTCTGTCAGTGGCTCGCGCCTCGCGGACTCGCAGCACTGGCAC 16054  
 QY 463 LeuGlu-----464  
 Db 16053 CTGAAGCGCGCGTCTGGCTGCTTAAGCATAGTCTTTCGCCCCACGCGTTTGGCTCTCC 15994  
 QY 465-----LeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThr 479  
 Db 15993 GTAGCAAGCTCTTGCTGCTCGCATAGCGGAGTAGCCGGTGGCGCTGGCGATCGCTC 15934  
 QY 480 ValCysCys-----GluGlyArgGluLeuLeuLeuCys-SerAsnThr 493  
 Db 15933 TC CGCTGCTCTTC CGCGAGCATAGGCGCGCGCTGCTACTACGCTGTCAATACTG 15874  
 QY 493 rSerCysCysArgCysPheCysValGluCysLeuGluValLeuValGly-----509  
 Db 15873 CATCCACAGTAGATGTAGTGCAGTAGGAGTATAAATACGAGCATCGGCCACCACTGCT 15814  
 QY 510-----ThrGlyThrAlaAlaGluAlaLysLeuGln--GluProThrS 523  
 Db 15813 CGATCATCATCGACGCTGTTTCTGTGGCGATACGCCCTTTAGGCGCCCTCATGGAG 15754  
 QY 523 rCysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArg-----538  
 Db 15753 CGCGAA-----GTCTCGCGCACTCGCACAGGGTCTGTGTAGAGCGTTTCGAGCAC 15700  
 QY 539-----ArgLysAspTyrAsnValArgLeuGlnAla-----548  
 Db 15699 CTCGCTACATCTTGTCTGG-----CGCGCGAGGCCCAACCGGTGTATTACTAGGTG 15646  
 QY 549-----PhePheThrSerAspThrGlyLeuGluTyrGluAlaProly 562  
 Db 15645 AGATAGAATGGACATTTTTTTTTTTTTTTAGAAAAGTCGGCTTTGAAAGGACGCGCGCGG 15586  
 QY 562 LeuTyr-ProAlaIleProAlaAlaArgArg 572  
 Db 15585 ACTATGCGCCAGG-----CCTTGTAGACGTAGG 15558

RESULT 10  
US-08-760-745-4  
; Sequence 4, Application US/08760745  
; Patent No. 5972658  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; APPLICANT: Murry, Lynn E.  
; TITLE OF INVENTION: LUNG GROWTH FACTOR VARIANT



ADDRESS: Dow AgroSciences LLC Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: USA  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/036,987A  
FILING DATE: 09-MAR-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stuart, Donald R  
REGISTRATION NUMBER: 28,479  
REFERENCE/DOCKET NUMBER: 50,608  
TELEPHONE: (317)337-4816  
TELEFAX: (317)337-4847  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80161 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-036-987A-1

Alignment Scores:  
Pred. No.: 10.6 Length: 80161  
Score: 129.50 Matches: 137  
Percent Similarity: 32.91% Conservative: 72  
Best Local Similarity: 21.57% Mismatches: 248  
Query Match: 2.84% Indels: 178  
DB: 30 Gaps: 30

US-09-720-086-8 (1-853) x US-09-036-987A-1 (1-80161)

QY	20	SerLeuValAsnGlyAlaCysSerAspGlnSerSerAspSerProPheLeuGlu	39
DB	29836	ACCGTCTGATTACCGCACTACCAACGACACCGCCAGCCTCATGCCCTCCGCG	29777
QY	40	AlaIleArgThrPro-----GluIle	46
DB	29776	GGCATCCGACACCGCTCCAGCAACACAGTCCGGCACCTCACCCATCCAGTGCATC	29717
QY	47	ArgGlyArgArgSerSerArgLeuSerLysArgGluValSerSerLeuSerTyr	66
DB	29716	CGACCCCGCGGACGACGACTTGCACCCCATCCGGCGGAGACCCCGTGGCGGGAGAA	29657
QY	67	ThrGlnAspLeuThrGlyAspGlyAspGlyGluAspGlyAspGlySerAspThrProVal	86
DB	29656	CTCCAGGAACACCTTCGCGCGTCGACATCACCGGTGACGCCACCGCAAGGGCAAGTCCGA	29597
QY	87	MetPro-----	88
DB	29596	CTCACCGGCGCGAGTGTGTTCCCGCCAGGTGCAACGCCACCAACGACGACGACACGCG	29537
QY	89	LysLeuPheArgGluThrArgThr-----ArgSerGluSerProAla-----	102
DB	29536	CGTATCCACCGTCAACCGGAGACCTCGAACCCGAAACGAATACGCAACCCCGGCCGAAAG	29477
QY	103	-----ValArgThrArgAsnAsnAsnValSerSerArgGluArgHisArgProSer	120
DB	29476	AACCGCTCCGCGACCGCCCTCAGCACATGACCTGCCACCGCGCGCGCGCTCCGCCAA	29417
QY	121	ProArg-----SerThrArgGlyArgGlnGly	129
DB	29416	CCACGACGCGTAGTCTCGGCCATTGGTGGCGAGAACACGCGCAACGCGGGAACCGCGTGT	29357

QY	130	ArgAsnHisValAspGluSerProValGlu---PhePro-AlaThrArgSerLeuArgAr	148
DB	29356	CGACCGCTGTCTCCAGCCAGCCCGCTCGAGAGACCTCCACACGACGACTCCACGACGACCG	29297
QY	148	g-----ArgAlaThrAlaSerAlaGly-----	155
DB	29296	CTGCTGGGATCCATCGCAGCGCTTCACGCGGCGAATCCCGAGAAACCCCGCTCGAA	29237
QY	156	-----ThrProTrpProSerProSerSerTyrLeuThrIle-----	168
DB	29236	CTCCGACACGCGCTTGAAGAAACCCCGCGGAGCATACGAGTCCGAGAGCATCGGA	29177
QY	169	AspLeuThrAspAspThrGluAspThrHis-----GlyThrProGlnSerSe	184
DB	29176	GTCCGGAACGAACACCCCATCCAGGTCCCATCCCGGACCTCCCGGAGATCCG	29117
QY	184	serThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyGlyMetGluSer	202
DB	29116	GTCCACACCGCGCGCA-----GCCAGTTGCCACAGGTCTCCGCGGAGGAGACCTGACC	29063
QY	203	-----ProGlnValGluAlaAspSerGlyAspGlyAspSerSerGluTyrGlnAsp--	219
DB	29062	GGGATAACGACAACTCATCGCCACGATCGGATGGATCGCTCTCTCCCGGAGAACCG	29003
QY	220	-----GlyLysGluPheGlyIleGlyAsp-LeuValTrpGlyLysIleL	234
DB	29002	ATCCCTCGCGCGCGGAGGGTCTGTTCTCTTTCGAGCGATGCTCGCAGTGCCTCGACAC	28943
QY	234	ysGlyPheSerTrpTrpProAlaMetValSerTrpLysAlaThrSerLysArgGlnA	254
DB	28942	TTCTTCGTAACTGG-----TGGTCACTCATCTCCAGGTCCG	28904
QY	254	laMetSerGlyMetArgTrpValGlnTrpPheGlyAsp-GlyLysPheSerGluValSer	273
DB	28903	CGGTCTCAGGACTCGGCTCGCCACGCGCTCTTCGACGAGCGCGCTACATCCATCGCA	28844
QY	274	AlaAspLysValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThr-----	291
DB	28843	TCGATCGAGGTCCGCGCGCATCG-----TCCGAGGAGATACCGAGTCCGCGGTGT	28790
QY	292	-----PheAsnLysLeuValSerTyrArgLysAla-----	301
DB	28789	CGCGCAAGTCGACGACGCGACTCCACCAAGCTCGCTCG--CGCAAGCGCGGAACGG	28733
QY	302	-----MetTyrHisAlaLeuGlyLysAlaArgValArgAlaGlyLysThrPheProSer	319
DB	28732	ATCGACGTCAGTCCGCTCTGATTCCGCGCTCGTCTCTCTCGGGAATGCTCTCTCGG	28673
QY	320	SerProGlyAspSerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGly	339
DB	28672	GACTCGGAGAACGCTCGCC-----CGCAGCTCCGCA	28640
QY	340	PheLysProThrGlyIleGlyLeuLys-----ProAsnAsnThrGlnProVal	356
DB	28639	GCCAGTTCGCGCGCGTGGATGGCTGAAGATAATTGCGATGGCAACTGGAGCGCGGTA	28580
QY	357	ValAsnLysSerLysValArgAlaGlySerArgLysLeuGluSerArgLysTyrGlu	376
DB	28579	GCTCGATTGACCGCAACG--CGAGTTCGACCGCGCGCTCGAGTCCGATCGCTGTC	28523
QY	377	AsnLysThrArgArgThrAlaAspAspSerAlaThrSerAspTyrCysProAlaPro	396
DB	28522	TTGAAGGCTCGT-----TCCGATCAACGCGCTCGCGGGGTGTTCCAGCACC	28472
QY	397	Lys-----ArgLeuLysThrAsnCysTyrAsnAsnGly---	407
DB	28471	AGACGGGTTTCGCGCGTACCAGTCCAGGATGCTCTCTGTGTCGCGCTCGGTGCC	28412
QY	408	-----LysAspArgGlyAspGlnSerArgGlnMetAlaSerAspVal	424
DB	28411	CGGACAAATCGTCCGCCATTCCAACGCGCTCGCAAGGACTCCCGCTCTCGGGTGG	28352
QY	425	AlaAsnAsnLysSerSerLeuGluAsp---GlyCysLeuSerCysGlyArgLysAsnPro	443





```
QY 357 ValAsnLysSerLysValArgAlaGlySerArgLysLeuGluSerArgLysTyrGlu 376
Db 28579 GCTCGATTGAGCCGAACG---CGAGTTCCAGACCGCGCTCGGAGTCAATCCCGCTGCC 28523
QY 377 AsnLysThrArgArgThrAlaAspAspSerAlaThrSerAspTyrCysProAlaPro 396
Db 28522 TTCAGAGCTCGT-----TCCGATGCAACGCGCGCTGCCCGGGGTGTCCAGCAC 28472
QY 397 Lys-----ArgLeuLysThrAsnCysTyrAsnAsnGly--- 407
Db 28471 AGAGCGGTTTCGCGGTACCAAGCTCCAGGATCAGCTGTCTGTGTCGGCTCGGTGCC 28412
QY 408 -----LysAspArgGlyAspGluAspGlnSerArgGluGlnMetAlaSerAspVal 424
Db 28411 GCGGCAATCGCTCGCGCATTCACAGACGCGTCCGGCAGGACTCCCGCTCGGGCTGG 28352
QY 425 AlaAsnAsnLysSerSerLeuGluAsp---GlyCysLeuSerCysGlyArgLysAsnPro 443
Db 28351 GCATCCAGCAGGAGTTCCTTGACCTCGGCCAGTTCCTTCAGCAGCGCGCGC----- 28301
QY 444 ValSerPheHisProLeuPheGluGlyGlyLeuCysGlnThrCysArgAspArgPheLeu 463
Db 28300 -----GGCTAGCCGCCGAC----- 28286
QY 464 GluLeuPheTyrMetTyrAspAspGlyTyrGlnSerTyrCysThrValCysCysGlu 483
Db 28285 -----AATCCCGCGCTGAAACGTCGCCCAATCGACATCGGCCACT 28247
QY 484 GlyArgGluLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCysValGluCys 503
Db 28246 ACCAGGAAATTCGTCGTGTC-----ACGGCTCGCGGATTCGGCGATGGCGTGC 28193
QY 504 LeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnGluProTyrSer 523
Db 28192 -----TGGGAGCCATTGGGGAAAGCCGATATCGCGCGCAGCTGTCT 28151
QY 524 CysTyrMet-----CysLeuProGlnArgCysHisGly 534
Db 28150 CTGTTTCTCCGGAGGCGCATGCTCTCCCGGCCACGCG 28112
```

## RESULT 13

```
US-09-491-772-1
; Sequence 1, Application US/09491772
; Patent No. 6329172
; GENERAL INFORMATION:
; APPLICANT: Rhee, Joon Shick
; APPLICANT: Pan, Jae Gu
; APPLICANT: Ahn, Jung Hoon
; TITLE OF INVENTION: ABC Transporter Gene Cluster in Pseudomonas fluorescens for
; FILE OF INVENTION: Enhanced Lipase Secretion
; FILE REFERENCE: 0136/0650
; CURRENT APPLICATION NUMBER: US/09/491,772
; PRIOR FILING DATE: 2000-01-26
; PRIOR FILING DATE: 1999-03-13
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: KOPATIN 1.5
; SEQ ID NO 1
; LENGTH: 8580
; TYPE: DNA
; ORGANISM: Pseudomonas fluorescens
US-09-491-772-1
```

```
Alignment Scores:
Pred. No.: 0.535 Length: 8580
Score: 126.00 Matches: 104
Percent Similarity: 34.84% Conservative: 42
Best Local Similarity: 24.82% Mismatches: 139
Query Match: 2.76% Indels: 137
DB: 4 Gaps: 16
```

```
US-09-720-086-8 (1-853) x US-09-491-772-1 (1-8580)
QY 85 ProValMetProLysLeuPheArgGluThrArgThrArgSerGluSerProAlaValArg 104
Db 5708 CTTATTACACCCAGCGCTGTTTG---CCAGGACCATGATCGATCGCCGCGCAAGA 5764
QY 105 ThrArgAsnAsnAsnSerValSerSerArg-----GluArgHisArg 118
Db 5765 AGAAGGCTTCGAACAGCAGTTCACAGACAGCCGCACTGTTCCAGCAGGCGGAGGCA 5824
QY 119 ProSerPro-----ArgSer 123
Db 5825 CCGCACCGATATTCGAGCGAGTTCGCTTATGACTGCGCACCGCGCGAGGATCG 5884
QY 124 ThrArgGlyArgGlnArgAsnHisValAspGluSerProValGluPheProAlaThr 143
Db 5885 AGGCGCTGACAGACAGCAGCGCGCTCGCGAGCTGGGGCGCTGATCGGAGTACAGA 5944
QY 143 rArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProTyrProSerProProSe 163
Db 5945 GCGTCAACATCGACGACTGGCGCGCTGAGTCC-----TGGCT-TCCGCGCTTC 5994
QY 163 rSer-----TyrLeuThr-IleAspLeuThrAspAspThrGluAspThrHisG 179
Db 5995 AGCTTGAGCCCGGCAACTACGACACCTGGCAGCGTGGCGATCAGCAATAACCCAGC 6054
QY 179 ly-ThrProGlnSerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGly 198
Db 6055 CTGGCATCCAGCGCCAGGCGCTGGAGTGGCGGCTATGAAGTGGAGCGCAACCGCGC 6114
QY 199 Gly-----MetGluSerProGlnValGluAlaAspSerGly 210
Db 6115 GGGACCTGCGGAAGTCAACCGCTATGCGAGTTCGCGCCAGCAGGAGTTCGACACGCGC 6174
QY 211 AspGlyAspSerSerGluTyrGlnAspGlyLysGluPheGlyIleGlyAspLeuValTyr 230
Db 6175 RACACTACACACGCGCTAT---GACACCAACACCTCGCGCTCGAGTCAACCTG--- 6228
QY 231 GlyLysIleLysGlyPheSerThrProAlaMetValSerThrTyrLysAlaThrSer 250
Db 6229 -----CCGTTGTATCCGCTGGCGCTCTCGCGCTCC 6261
QY 251 LysArgGlnAlaMetSerGlyMetArgTyrValGlnThrPheGlyAspGlyLysPheSer 270
Db 6262 ACTCCAGCGCCAGCGCGCTGAGTTCGCGCGAAGCGTTGGTGTCTCTACCGCGCAAGC 6321
QY 271 Glu-----ValSerAlaAspLysLeuValAla----- 271
Db 6322 GAGACCTTGTATCGAACTGCGTCCCAATTCAGCGCGTCTTGTCCGGGTGAGCAAGTTG 6381
QY 272 -----ValSerAlaAspLysLeuValAla----- 279
Db 6382 CGCGGTACACAGAGCGCGTGCCTGCGCGAAGCGTTGGTGTCTCTACCGCGCAAGC 6441
QY 280 ---LeuGly-----LeuPheSer 284
Db 6442 ATCTTGGCGGTGAGCGGTCAATCTTGATGCGTTGAATCCGAGCAGCAGCTGTACAGC 6501
QY 285 GlnHisPheAsnLeu-----AlaThrPheAsnLysLeuValSerTyrArgLysAlaMet 302
Db 6502 ACACCGCAGACCTTCGCGCGCGTGTGACTTACTTGTGCGCTGGACCAAGTGTGAT 6561
QY 303 TyrHisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSerProGly 322
Db 6562 TACTACGCGCG---CAATTTGCGGACACCGA-----CCTGCCAAGGTAGA 6605
QY 323 AspSerLeuGluAspGlnLeuLysProMetLeuGluTyrAlaHisGlyGlyPheLysPro 342
Db 6606 CGAAGCCTTCGGGACCAAGA-GGGCGAGCCTC----- 6637
QY 343 ThrGlyIleGluGlyLeuLysProAsnThrGlnProValValAsnLysSerLysVal 362
Db 6638 -----CGCTGCAGACAGCGCCCTGTTGCAACAAACAAACCGCCTG 6676
```

QY 363 Arg-----ArgAlaGlySerArgIleuGluSerArgIleuGlu 376  
 Db 6677 AGGCAAGCCGACCCCTGTGTGCGAGCGGGCTTGCCCGCTTGCGCAACACCC 6736  
 QY 377 AsnIleThrArgAlaGlyThrAlaAspAspSerAlaThrSerAspIleuGluPro 396  
 Db 6737 AATTAACCCACACATCTCCGCGCAATTCAC-----AAC 6772  
 QY 397 LysArgIleuThrAsnIleuGluSerAspAlaGlyAspGlu 413  
 Db 6773 AAGGAGACAGACAGATGGGTATTGACTACAGAACCTCGGACCGAA 6823  
 RESULT 14  
 US-09-179-558-54  
 Sequence 54, Application US/09179558  
 Patent No. 6180612  
 GENERAL INFORMATION:  
 APPLICANT: Hockensmith, Joel W.  
 APPLICANT: Muthuswami, Rohini  
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
 TARGETING DNA METABOLIC PROCESSES USING  
 AMINOGLYCOSIDE DERIVATIVES  
 NUMBER OF SEQUENCES: 66  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: PENNIE & EDMONDS LLP  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10036-2711  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/179,558  
 FILING DATE: 27-OCT-1998  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: U.S. 09/060,470  
 FILING DATE: 15-APR-1998  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: U.S. 60/063,898  
 FILING DATE: 31-OCT-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Coruzzi, Laura A  
 REGISTRATION NUMBER: 30,742  
 REFERENCE/DOCKET NUMBER: 9426-005-999  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212)7909090  
 TELEFAX: (212)8699741  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 54:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2874 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: Other  
 US-09-179-558-54

## Alignment Scores:

Pred. No.: 0.0951 Length: 2874  
 Score: 125.50 Matches: 195  
 Percent Similarity: 35.60% Conservative: 130  
 Best Local Similarity: 21.36% Mismatches: 329  
 Query Match: 2.75% Indels: 262  
 DB: 4 Gaps: 55

US-09-720-086-8 (1-853) X US-09-179-558-54 (1-2874)

QY 35 ProProIleIleuGluAlaIleArgThrProGluIleArgGlyArgGlySerSerArg 54  
 Db 25 CCTTGCTCTTACAGAGAGAGAGAAATAATTAAGCAATTCGACAGAGCT--- 81  
 QY 55 LeuSerIleThrArgIleuValSerSerIleuLeuSerIleuThrGluIleuValAspGly 74  
 Db 82 CTGGCCCGAAGAGCTGAGAAACTATTATGACAGAACAGATCAGAAA-----CTGCC 132  
 QY 75 AspGlyIleuAspGlyAspGlySerAspThrProValIleProIleuPheArgGluThr 94  
 Db 133 CAGTCACAGACAGGCGCCATCCCAAAACCTCCCGAGATCT----- 174  
 QY 95 ArgThrArgSerGluSerProAlaValArgThrArgAsnAsn---SerValSerSer 113  
 Db 175 ---TCAAGTCAGGAGCCATGCGATCTTTTCAACAAATAATCCAGAGTTCACT 231  
 QY 114 ArgIleuArgIleuArgProSer-----ProAspSerThrArgGlyArg 127  
 Db 232 CATGTGACACAGACAGCTCAAAATCCCAAGTTTCCACCCCAACCTCGACAGCG 291  
 QY 128 GlnGlyArg---AsnIleValAspGlySerProValGluPheProAlaThrArgSerIleu 146  
 Db 292 AAGGAGATGCGCAGAGCCAGAGAGATGCCACAGCTGCCAAGCTACCGCCACCA 351  
 QY 147 ArgArgAlaThrAlaSerAlaGlyThrProTrp-----ProSerProPro 162  
 Db 352 AATCAAGTACGTGCTGCGGATCTCCCTGCCCTGCAAAACAGCTCTCGAGGATCCC 411  
 QY 163 SerSerIleuThrIleuAspIleuThrAspAspThrGluAspThrIleuGlyThrProGln 182  
 Db 412 AGCCACAGCTTTGG-----GGTTGTAGTTAGGTCAAGGTCATCTCAG 456  
 QY 183 SerSer-----SerThrProTrpAla-----ArgIleu 191  
 Db 457 GCTTCACCTCAGACCCAGCTACACCTTCCTCAACACTCAGACGCTTGGCGCAA 516  
 QY 192 AlaGlnAspSerGlnGlnGlyIleuGlySer-----ProGln 204  
 Db 517 GTGAAGATTTCAGAGACAGACAGAGCTCTTCGTGACAGCTCTTCAAGGATCTGAA 576  
 QY 205 ValGluAlaAspSerGlyAspGlyAspSerSerGluIleuValGluGluPheGly 224  
 Db 577 TTAGAGCCAGATCGAAGACCTTCACCTCT-----GGGAG-----AAC 618  
 QY 225 IleGlyAspIleuValIleGlyIleuValIleuGlyPheSerTrpTrpProAlaMetValVal 244  
 Db 619 ATTCCGGGAGTGTGATGCCAGACAGAAAGCA-----AGACTG 657  
 QY 245 SerTrpIleuValThrSerIleuValAlaMetSerGlyMetArgTrpValGlnTrpPhe 264  
 Db 658 CACAGAAAGCAGGAGCCCGATGACAGAGTGTGAGCTCCAGAGGAGAGGTATC 717  
 QY 265 GlyAspGlyIleuPheSerGluValSerAla-----AspIleuValAlaIleu--- 280  
 Db 718 CGAAGCAGAGAGCATTCACAGTTCAGATTCGATTCATGAGGCGCTCATGCGAGTGT 777  
 QY 281 ---GlyLeuPheSerGlnIleuPheAsnIleuAlaThrPheAsnIleuValSerIleuArg 299  
 Db 778 AAGAGTCTGCCAGAGAGATGATGATCTCTCAC-----AAGACGTGAATCTTACG 831  
 QY 300 LysAlaMetIleuValAlaIleuGluValAspAlaArgValAlaArgAlaGlyIleuThrPheProSer 319  
 Db 832 ATGACTGACTATGTGCTCCCTAATGAAGCAGCCAGCGCTC----- 873  
 QY 320 SerProGlyAspSerIleuGluAspGluIleuValProMetIleuGluTrpAlaHisGlyGly 339  
 Db 874 ---CCAGGATCACCCTCGACCT-----TTGGAAGAGCGGAGGCGCAC 915  
 QY 340 PheIleuProThrGlyIleuGluGlyIleuValProAsnIleuThrGlnProValValAsnIleu 359  
 Db 916 ATGGAG-----TCACCTTCACACAGAGTGGCATTAATAGCCAA 954  
 QY 360 SerIleuValArgArgAlaGlySer----- 367



; INFORMATION FOR SEQ ID NO: 29:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17612 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-911-853-29  
  
Alignment Scores:  
Pred. No.: 2 Length: 17612  
Score: 125.50 Matches: 140  
Percent Similarity: 30.46% Conservatave: 58  
Best Local Similarity: 21.54% Mismatches: 233  
Query Match: 2.75% Indels: 222  
Gaps: 31  
US-09-720-086-8 (1-853) x US-08-911-853-29 (1-17612)  
  
Qy 6 ArgHISLeuAenGlyGlu-----GluAspAlaGlyGlyArg 17  
Db 13086 CGGCACGCGCGGCGGGAAGAACTGGAAACGCCGAGGCTACCTGAAGAAAGATCCGGTCGA 13145  
  
Qy 18 GluAspSerIleLeuValAsnGlyAlaCysSerAspGlnSerSerAspSerProIle 37  
Db 13146 CCCCTGGGGCAACCAAGTACCTGCTGCGGGGCAACCGCGCAAGATCAGACTGTA 13205  
  
Qy 38 LeuGluAlaIleArgThrProGluIleArgGlyArgArgSerSerArgLeuSerlys 57  
Db 13206 TTCGCTGGGCGCCGACGGCCA-----GGAAAGCGGCGGAGGAGACCGACCGCACT 13256  
  
Qy 58 ArgGluValSerSerLeu-----LeuSerTyrThrGlnAspLeuThr 71  
Db 13257 CGGCACACTGGGATCTCTACTCGCAATGACAGCGGGGCGCGGTTCACCTGTGATGAGCT 13316  
  
Qy 72 GlyAspGlyAspGlyGluAspGlyAspGlySerAspThrProValMetProLysLeuThe 91  
Db 13317 GCTGGTGGTGGTGGTGGTGGTGG----- 13340  
  
Qy 92 ArgGluThrArgThrArgSerGlnSerProAlaValArgThrArgAsnAsnSerVal 111  
Db 13341 -----CGTGCACACGGGCTCGC----- 13358  
  
Qy 112 SerSerArgGluArgHisArgProSerProArgSerThrArgGlyArgGlnGlyArgAsn 131  
Db 13359 CGTGTGCGGACGGGATCGGCACAGCCCGCGCAAGCTGGCGAGCGAGCGCA--- 13415  
  
Qy 132 HisValAspGlnSerProValGlnPheProAlaThrArgSerLeuArgArgAlaThr 151  
Db 13416 -----GCGCCTGCAGTGGCTGCGCGGAGTGTGCTGACGA 13451  
  
Qy 152 AlaSerAlaGlyThrPro-----Trp-----ProSerProProSerSer 164  
Db 13452 GCGCGTGTGGAACAACCGCAGATATGGCTACGCTTCAACGCCCGGAGCTACCGGCTGCT 13511  
  
Qy 165 TyrLeuThrIleAspLeuThrAspAspThrGluAspThrHisGlyThrProGlnSerSer 184  
Db 13512 GCGCTT-----CGAGCCGCGACAGCGGCGCTGGAGACCGCT 13547  
  
Qy 185 SerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMetGlnSerProGln 204  
Db 13548 CGACGAGCGCGTGCAGAGCTGCC-----GGAGTG----- 13577  
  
Qy 205 ValGluAlaAspSerGlyAsp---GlyAspSerSerGluTyrGlnAspGlyLysGluPhe 223  
Db 13578 -----GCTCAGACTGGAGATCGAGGTGACAGCAGAGTGTGCGGCTGCGCGCGC--- 13628  
  
Qy 224 GlyIleGlyAspLeuValTyr-----GlyLysIleLysGlyPheSerTyrTyrProAla 241  
Db 13629 -----CCGTGGGAGACAGACAAAGCCGCGGCGCAAGGCGCACACTGCT 13673  
  
Qy 242 MetValValSerTyrLysAlaThrSerLysArgGlnAlaMet----- 255  
Db 13674 GCTGCTTCAGTGGCGAGCTGACCCCTTGCCTGTGCGCTGTGCGCGCGCGAGCG 13733

Qy 256 -----SerGlyMetArgTyrValGlnTyrPheGlyAspGlyLysPhe 269  
Db 13734 CGGCGCGCGGTGCTGACGCTGGCCAGCGACGGCTT-----CGCCGA 13775  
  
Qy 270 SerGluValSerAlaAspLys----- 276  
Db 13776 GCGCGAGCTGCAGAGGAAAGATCCCGATGAAGCGCGCGCTTACCTGCTGCAG 13835  
  
Qy 277 ---LeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsnLysLeu 295  
Db 13836 GTGCTGGTGGCCCTGGCGATGCTTCC----- 13862  
  
Qy 296 ValSerTyrArgLysAlaMetTyrHisAlaLeuGlnLysAlaArgValArgAlaGlyLys 315  
Db 13863 -----GTGTCGCGCGCACGCTGCTCAGCGCCAGCGGCTGCTCGCTGAG 13907  
  
Qy 316 ThrPheProSerSerProGlyAspSerLeuGlnAspGlnLeuLysProAsnAsnThrGln 335  
Db 13908 ACC-----GCCGCGCGCTGAGGAC-----AAGACCTTCCGACACTGG 13946  
  
Qy 336 AlaHisGlyGlyPheLysProThrGly---IleGluGlyLeuLysProAsnAsnThrGln 354  
Db 13947 CT---GGCGACAAACGCTTCAGAGACTGACAGCTGACGAGTCCGCGCGGAGCGG 14002  
  
Qy 355 ProValValaAsnLysSerLysValArgArgAlaGlySerArgLysLeuGlnSerArgLys 374  
Db 14003 CCGCGAGCGAGCGCGAGAGACTACCGCGGCGCGCTGCTGGCGAG--- 14053  
  
Qy 375 TyrGluAsnLysThrArgArgThrAlaAspSerAlaThrSerAspTyrCysPro 394  
Db 14054 -----CGAGTGCAGCGCACACGACCGGAGATGTC-TGCCGTG 14091  
  
Qy 395 AlaProLysArgLeuLysThrAsnCysTyrTrpAsnAsnGlyLysAspArgGlyAspGlnAsp 414  
Db 14092 TCACCGTACGGG-----TGCGCTGCGGCGCGGACCGCGGCTGCAGGCGCA 14136  
  
Qy 415 GlnSerArgGluGln-----MetAlaSerAspValAlaAsnAsnLysSer 429  
Db 14137 AGATCGAAGACCAACCCCTGCTGACCCGATGAGTCTTGTGGGGTGCAGCAGTGAAGC- 14195  
  
Qy 430 SerLeuGluAspGlyCysLeuSerCysGlyArgLysAsnProValSerPheHisProLeu 449  
Db 14196 AGCGCGCTTCAACCTGCTGGAAGTGCATGACCATGCGCATCTTGCCTGTGGCCA 14255  
  
Qy 450 PheGluGlyGlyLeuCysGlnThr-----CysArgAspArgPheLeuGlnLeuPheYtr 467  
Db 14256 TGGCCACTTACCGCATGCTGCAGACCGTGTGCAGACCGATCGTGGCCAGC----- 14306  
  
Qy 468 MetTyrAspAspAspGlyTyrGlnSerTyrCysThrValCysCysGlnGlyArgGluLeu 487  
Db 14307 -----GCCAGCAGAGACAGCGTCTCGGAGGTGACGCGGCGCA 14345  
  
Qy 488 LeuLeuCysSerAsnThrSerCysAspArgCysPheCysValGluCysLeuGluValLeu 507  
Db 14346 TGGCAGCTTTCGAACGCGACCTGCTGCAGTGGCGCTGCGTCCGCTG- 14393  
  
Qy 508 Val-GlyThrGlyThrAlaAlaGluAlaLysLeuGlnGluProTyrPheSerCysTyrMetCys 527  
Db 14394 GCGACCGCGTGGGAGACCTGTGCGACCGCTGCAGCGGAGAGAGTGG----- 14439  
  
Qy 527 sLeuProGlnArgCysHisGlyValLeuArgArgArgLysAspTyrPheAsnValArgLeuGln 547  
Db 14440 -----CCGGAACACCAAGTGAAGTTCACCGCGAGCG----- 14472  
  
Qy 547 nAlaPhePheThrSerAspThrGlyLeuGluTyrGluAlaProLysLeuTyrProAlaIle 567  
Db 14473 -----CTGGCGCAACCGGCTCG 14489  
  
Qy 567 eProAlaAlaArg-----ArgArgProIleArgValLe 578  
Db 14490 GCCAGCGCGCGCACCTTACAGCGGAGTGGCTGCGAGCTTGAAGGAGCGGCTGGAGC 14549

Mon Nov 25 08:18:41 2002

us-09-720-086-8.rni

Page 18.

Qy 578 uSerLeupheAspGlyIleAlaThrGly 587  
Db 14550 GCGTTACTGGACGGTGTGCTGGACCAAGC 14577

Search completed: November 25, 2002, 03:06:35  
Job time : 473.863 secs



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

## OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 24, 2002, 12:10:37 ; Search time 72.5572 Seconds  
(without alignments)  
3854.741 Million cell updates/sec

Title: US-09-720-086-7  
Perfect score: 4935  
Sequence: 1 MPMPSSGPGDTSSTSSAARE.....MSVPVIRHLPAPLKEYFACV 912

Scoring table:  
BLOSUM62 Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 44162 seqs, 15333831 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09720086/runat.18112002.092957.29457/app.query.fasta\_1.4252  
-DB=Ibsed.Patentc.NA -QMT=faetap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsme2 -TRANS=numan40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09720086 @CGN 1.108 @runat.18112002.092957.29457 -NCPU=6 -ICPU=3  
-NO XLPVY -NO MMAP -LARGJURY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-MARK TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELXT=7

## Database:

1: Issued Patents NA: \*  
2: /cgn2\_6/ptodata/1/ina/5A.COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/5B.COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6A.COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/6B.COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	748	15.2	2077	4	US-09-276-531-47 Sequence 47, Appl
2	182.5	3.7	15872	4	US-09-105-537-1 Sequence 1, Appl
3	182.5	3.7	43280	2	US-08-804-227C-1 Sequence 1, Appl
4	172.5	3.5	12001	1	US-08-458-568A-11 Sequence 11, Appl
5	171	3.5	4852	1	US-07-853-913-3 Sequence 3, Appl
6	166	3.4	3014	1	US-08-629-939-1 Sequence 1, Appl
7	166	3.4	3014	1	US-08-759-873-1 Sequence 1, Appl
8	166	3.4	3251	4	US-09-085-199B-6 Sequence 6, Appl
9	166	3.4	35060	3	US-08-814-095-7 Sequence 7, Appl
10	163.5	3.3	4350	4	US-09-295-553-37 Sequence 37, Appl
11	161	3.3	7791	3	US-08-949-386-23 Sequence 23, Appl
12	161	3.3	7791	3	US-08-450-562-23 Sequence 23, Appl

13	161	3.3	7791	4	US-08-984-709A-23 Sequence 23, Appl
14	161	3.3	7791	4	US-08-490-272-23 Sequence 23, Appl
15	161	3.3	7808	3	US-08-949-386-22 Sequence 22, Appl
16	161	3.3	7808	3	US-08-450-562-22 Sequence 22, Appl
17	161	3.3	7808	4	US-08-984-709A-22 Sequence 22, Appl
18	161	3.3	7808	4	US-08-450-272-22 Sequence 22, Appl
19	159	3.2	2277	1	US-08-676-567-5 Sequence 5, Appl
20	159	3.2	2277	1	US-08-676-974-5 Sequence 5, Appl
21	159	3.2	2277	2	US-09-098-487-1 Sequence 1, Appl
22	159	3.2	8488	1	US-07-945-283-1 Sequence 1, Appl
23	158.5	3.2	2680	4	US-09-063-035-1 Sequence 1, Appl
24	158.5	3.2	9551	1	US-08-056-200-93 Sequence 93, Appl
25	158.5	3.2	9551	2	US-08-800-644-93 Sequence 93, Appl
26	157	3.2	20235	1	US-07-642-734C-3 Sequence 3, Appl
27	157	3.2	20235	3	US-08-439-009A-3 Sequence 3, Appl
28	156.5	3.2	1736	4	US-09-162-524-2 Sequence 2, Appl
29	156.5	3.2	7808	2	US-08-149-097D-22 Sequence 22, Appl
30	155	3.1	1681	4	US-09-434-288-7 Sequence 7, Appl
31	155	3.1	5468	4	US-09-535-008-66 Sequence 66, Appl
32	155	3.1	5471	4	US-09-535-008-1 Sequence 1, Appl
33	155	3.1	5477	4	US-09-535-008-74 Sequence 74, Appl
34	155	3.1	5480	4	US-09-535-008-70 Sequence 70, Appl
35	155	3.1	5564	4	US-09-535-008-68 Sequence 68, Appl
36	155	3.1	5567	4	US-09-535-008-64 Sequence 64, Appl
37	155	3.1	5573	4	US-09-535-008-76 Sequence 76, Appl
38	155	3.1	5576	4	US-09-535-008-76 Sequence 76, Appl
39	155	3.1	13842	4	US-09-105-537-30 Sequence 30, Appl
40	155	3.1	36778	4	US-09-105-537-7 Sequence 5, Appl
41	155	3.1	38506	3	US-09-320-878-19 Sequence 19, Appl
42	154	3.1	5467	1	US-07-745-206A-12 Sequence 12, Appl
43	154	3.1	5467	2	US-08-311-363-12 Sequence 12, Appl
44	154	3.1	11907	4	US-08-061-376-4 Sequence 4, Appl
45	154	3.1	14255	1	US-08-320-559-1 Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-09-276-531-47  
Sequence 47, Application US/09276531  
Patent No. 6183968  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Yue, Henry  
APPLICANT: Reddy, Roopa  
APPLICANT: Guebler, Karl J.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276,531  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/079,677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:









Db 7100 CGGGCGGACCGCCCAAGGGGGGGGGGGGGGGGGTAAAGAGAGTGAACCGAAGCG 7159

Qy 477 -----GluArgLeuValTyrGluValArgGlnLysCysArgAsn 489

Db 7160 TTCCGACTTCGTCCTCAATATATATATATTATTAGGCGGAAGTCCGAGCAC 7209

## RESULT 5

US-07-853-913-3

Sequence 3, Application US/07853913

Patent No. 5338839

GENERAL INFORMATION:

APPLICANT: McKay, Ronald D.G.

APPLICANT: Lendahl, Urban

TITLE OF INVENTION: Nestin Expression As An Indicator of

TITLE OF INVENTION: Neuroepithelial Tumors

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Militia Drive

CITY: Lexington

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02173

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/853,913

FILING DATE: 19920319

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/660,412

FILING DATE: 22-FEB-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/603,803

FILING DATE: 25-OCT-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/201,762

FILING DATE: 02-JUN-1988

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/180,548

ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227

REFERENCE/DOCKET NUMBER: MIT-4641AAAA

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-861-6240

TELEFAX: 617-861-9540

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 4852 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-07-853-913-3

## Alignment Scores:

Pred. No.: 0.00378

Score: 4852

Matches: 126

Conservative: 75

Mismatches: 191

Indels: 231

Gaps: 27

US-09-720-086-7 (1-912) x US-07-853-913-3 (1-4852)

Qy 6 SerSerGlyProGlyAspThrSerSerAlaAlaGluArgGluGluAspArgLysAsp 25

Db 2689 AGCATGGAACCTGGAGATTGGAGATCTCCAGAGGAGTAGACAAGGAAAGTCAAGGAAT 2748

Qy 26 GlyGluGluGlnGluGluProArgGlyLysGluGluArgGlnGluProSerThrThrAla 45

Db 2749 CTGGAAGAG---GAAGAGAACCTGGGAAAGGAGAGTAGTACCAAGAG---TCACGTG 2796

Qy 46 ArgLysValGlyArgProGlyArgLysArgLysHisProProValGluSerGlyAspThr 65

Db 2797 AGGTCTCTGGAGGAGGAGGAGGAGGAGTGGCG---CAGTCTGCAGATGTG 2844

Qy 66 ProLysAspProAlaValLysSerLysSerProSerMetAlaGlnAspSer---82

Db 2845 CAGAGGTGGGAGNAGTACGTTGGAGAGGAGGAGGAGTGGCTCAGGAAAGGCTCTCTGGG 2904

Qy 82 -----82

Db 2905 ATGGCTGGAGTGGGAAATAGGATGAGGAGAGCTGAATCTAAGGGAGGAGGATGGCTTC 2964

Qy 83 ---GlyAlaSerGluLeuLeuProAsnGlyAspLeuGluLysArgSerGlu---98

Db 2965 ACTGGGAAGGAGGAGGTGGTAGAGCAGGAGGAGGAGTGAATGCCAGAGGAGGTCTGGTTC 3024

Qy 99 -----ProGlnProGluGlu---103

Db 3025 CCAGGCGAGGGGACCCAGAGAACCTGTAGCCCAAGAGAGGAGGAGGCTGTGTGAGGGA 3084

Qy 104 GlySerProAlaGlyGlyGlnLysGlyAlaProAlaGluGlyGluGly---120

Db 3085 GCCAGTGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3144

Qy 121 -----AlaAlaGluThrLeuProGluAla---128

Db 3145 GGGACCCAGGCTCCAGGCTCCCGAGGGGCTGCCAGAGGCGATAGAGCCCTGTGGTGA 3204

Qy 128 -----128

Db 3205 GATGATGTGGCCCGAGGGGAGTACCAAGCTCCCGAGAGTCACTGTGGGGTCAGAGCCT 3264

Qy 129 -----SerArgAlaValGluAsnGlyCysCysThrProLysGluGlyArg 143

Db 3265 GCCATGGGTGAGTCTGCTCGGGAGCTGAGCCAGGCGCTGGGGGAGGGGTGGAGGGCTG 3324

Qy 144 GlyAlaProAlaGluAlaGlyLysGluGln---LysGluThrAsn 157

Db 3325 GGGGACCCAGGCCATCTTGACCAAGGAGGAGTGTGGAACCCACCCCTGGAAGAGGAGGT 3384

Qy 158 IleGluSerMetLysMetGluGlySerArgGlyArgLeuArg---Gly 172

Db 3385 TTGGAGGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3444

Qy 173 GlyLeuGlyTrpGluSerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAla 192

Db 3445 GGTCTGGGACAGAGTCTCC-----GAGCTGCTGGGAAGAGC 3483

Qy 193 GlyAspProTyrTrpLysSerLysArgAspGluTrpLeuAlaArgTrpLysArg 212

Db 3484 AGAGACCTTGGAGGCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3543

Qy 213 GluAlaGluLys-----LysAlaLysValLeuAlaGlyMetAsnAlaValGluGluAsn 230

Db 3544 GGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3603

Qy 231 GlnGlyProGlyGluSerGlnLysValGluGluAlaSerProProAlaValGlnGlnPro 250

Db 3604 CTTGGGCTCTGGGGTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3663

Qy 251 ThrAspProAlaSerProThrValAlaThrThrProGluProValGlySerAspAlaGly 270

Db 3664 -----AGCCCAACGTACACC-----3678

Qy 271 AspLysAsnAlaThrLysAlaGlyAspGluProGluTrpGluAspGlyArgGlyPhe 290

Db 3679 -----CCGATCTCGGAAGATGCCCTGGGCTC 3705

Qy 291 GlyIleGlyGluLeuValTrpGlyLysLeuArgGlyPheSerTrpProGlyArgIle 310



QY 238 LysValGluGluAlaSerProProAlaValGln-----Gln 249  
Db 560 -----CCGAGCCCTCCGCGCTGGAGGTCCACGAGCTTGGTGCTGTTTGGC 607  
QY 250 ProThrAspProAlaSerProThrValAlaThrThrProGluProValGlySerAspAla 269  
Db 608 CCGCAACTTCCCGAAGATCCACCGGCTGCCCGCCACCCAGCGGGTGTGTCCCGCTC 667  
QY 270 GlyAspLysAsnAlaThrLysAlaGlyAspAspGluProGluTyrGluAspGlyArgGly 289  
Db 668 ATGAGCGGCTCCGGTGCAGGTTGGAGACAGCTCC----- 703  
QY 290 PheGlyIleGlyGluLeuValTrpGlyLysLeu-----ArgGlyPheSerTrpTrpPro 307  
Db 704 -----GGACGGCAGCTGCCCATTAAGTCTCTCCCGGGGCTGTCA-----CCA 748  
QY 308 GlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGluGlyThrArgTrp 327  
Db 749 GCCCG----- 754  
QY 328 ValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLysLeuMetProLeu 347  
Db 755 -----CAGCTGCTGCTCCCGGCC 772  
QY 348 SerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnProMetTyrArgLys 367  
Db 773 TCT-----GAGAGCCCTCACTGCTCGGGGGCCCAAGTGAAGCGCTCTCCCGAGGCC 823  
QY 368 AlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPheProValCys 387  
Db 824 GCTCGCGTGGAGTTGAGGAG----- 844  
QY 388 HisAspSerAspGluSerAspThrAlaLysAlaValGluValGlnAsnLysProMetIle 407  
Db 845 GAGATGCTCTGAGTCTCGAGTCTGCGGTCCTCTGAAGGGCAAACTCG-- 901  
QY 408 GluTrpAlaLeuGlyGlyPheGlnProSerGly-----ProLysGly--- 421  
Db 902 -----GCTCTGGTGGCGCGCGCTGGAGGAGAGCGCGCTGCTCCCGGGGGCG 955  
QY 422 -----LeuGluProProGluGluGluLysAsnProTyrLysGluVal 435  
Db 956 GCACGAGAGCGCTCGCCCTGCTCCCAAGGAAGATTCCCGTCTTCAGCGCCCGAGGTC 1015  
QY 436 TyrThrAspMetTrpValGluProGluAlaAlaTyrAlaProProProAlaLys 455  
Db 1016 -----GCCCTGCTGGAGCAGGAC-----GCCCGATGGCGCGCGCGC 1054  
QY 456 LysProArgLysSerThrAlaGluLys-----ProLysValLys 468  
Db 1055 TCCCGCTGGCCACCGATGATGATTCATCCAGTGCCTATCTGCTCTCAATCAC 1114  
QY 469 GluIleAspGluArgThrArgGluArgLeuValTyrGlu--ValArgGlnLysCysA 488  
Db 1115 GCCTTATTGGCAGCGGCTGCGAGCTGCTGGAAGACAAAGTTACAGCGGGGGCC 1174  
QY 488 rgaenleGluAspIleCysIleSerCysGlySerLeuAsnValThrLeuGluHis-Pro 507  
Db 1175 GGGCTGCGCAGCGCTTTTC-----CCCGCGGAGTTACCTGCTGCTGCTCCACCGC 1231  
QY 508 LeuPheValGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGluCysAlaTyrGln 527  
Db 1232 GTCCGTGTAGGCGAC-----TTCCCGACTGCGGTACCCG 1267  
QY 528 TyrAsp-----AspAspGlyTyrGlnSerTyrCysThrIle----- 539  
Db 1268 CCGAGCGCGAGCCCAAGGAGCGAGCTGATACCTCTCTATAGGAGACTTCACCGCCCGCT 1327  
QY 540 -----CysCysGlyGlyArgGluVal 546  
Db 1328 CTAAGATAAAGGAGGAGGAGCGCGGAGGCTCCCGCGCTCCCGCGCTCTCTCTAC 1387

QY 547 LeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysValAspLeuLeu 566  
Db 1388 CTTGTGGCGGTGCCAAC----- 1405  
QY 567 ValGlyProGlyAlaAlaGlnAlaAlaLysLysGluAspPro-TrpAsnCysTyrMetCys 586  
Db 1406 -----CCCGCAGCTTCCCGGATTTCCTCGTTGG----- 1433  
QY 586 sGlyHisLysGlyThrTyrGlyLeuLeuArgArgGluAspTrpProSerArgLeuGln 606  
Db 1434 -GGCCACCGCCCGC-----TGCGCGCGCGAGCGCCCATCCA----- 1472  
QY 606 nMetPheAlaAsnHisAspGlnGluPheAspProProLys-----Va 622  
Db 1473 -----GACCGGGGAGCGCGGTGACGGCC 1498  
QY 622 ltyrProValProAlaGluLysArgLysProIleArg 635  
Db 1499 GCACCGCGCAGTCTCAGTCTGCTGCTCTCTCTCGG 1538  
RESULT 7  
US-08-759-873-1  
; Sequence 1, Application US/08759873  
; Patent No. 5683885  
; GENERAL INFORMATION:  
; APPLICANT: Kieback, Dirk G.  
; TITLE OF INVENTION: METHODS FOR DIAGNOSING AN INCREASED RISK  
; OF BREAST OR OVARIAN CANCER  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS  
; CITY: 2100 Pennsylvania Avenue, N.W., Suite 800  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08759.873  
; FILING DATE: 12-APRIL-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KIT, Gordon  
; REGISTRATION NUMBER: 30,764  
; REFERENCE/DOCKET NUMBER: A-6612  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 293-7060  
; TELEFAX: (202) 293-7860  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3014 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; HYPOTHETICAL: NO  
; US-08-759-873-1  
Alignment Scores:  
Pred. No.: 0.0046 Length: 3014  
Score: 166.00 Matches: 146  
Percent Similarity: 31.21% Conservative: 65  
Best Local Similarity: 21.60% Mismatches: 209  
Query Match: 3.36% Indels: 257  
DB: 1 Gaps: 31  
US-09-720-086-7 (1-912) x US-08-759-873-1 (1-3014)  
QY 53 ArgLysArgLysHisProValGluSerGlyAspThrProLysAspProAla----- 70



Db 9 CGCGCCCTCCCGCCCGCCGAGCCAGGA-GGTGAGATCTCCCGGTCCGACCAATTC 67  
 Qy 71 -----Val11eserlyserProserMetAla1asp 81  
 Db 68 AACACCACTTCTCCCTCCCTCCCTATATTCGCCGAACCCCTCTCTCTCCCTT 127  
 Qy 82 SerGlyAlaSerGluLeuLeuPro--AsnGlyAspLeuGlyAspSer----- 97  
 Db 128 TCC-----CTCTCTCTGAGAGACGGGGAGAGAGAAAGGGAGAGTCCAGTCGTC 175  
 Qy 98 -----GluProGln 100  
 Db 176 ATGACTGAGCTGAAGCAAGGATCCCGGGCTCCCACTGGGGGGCGCCCTCC 235  
 Qy 101 ProGluGluGlySerProAlaGlyGlyGlnGlyAlaProAlaGluGlyGly 120  
 Db 36 CCGAGAGTGGATCCCACTGCTGTGTCGCCAGCCGAGGTCCGTTCCGGGAGCCAG 295  
 Qy 121 AlaAlaGluThrLeuProGluAlaSerArgAlaValGlu-----AsnGlyCys 137  
 Db 296 ACCTCGACACTTGTCTGAAGTTTCG--GCCATCTATCTCCCTGAGAGGGCTACTC 352  
 Qy 138 ThrProlysgluGlyArgGlyAlaProAlaGluGlyGlnGlyGluThrAsn 157  
 Db 353 TTCCCTGGCCCTCCAGGAGCAGAGCCCTCCGACGAAAGACGACGACAGCTG 412  
 Qy 158 IleGluSerMetGluGlySerArgGlyArgLeuArgGlyGlyGlyGlyGly 177  
 Db 413 CTGTCGACGAGGGGGGCGCATTTCCAGAGCTGAGCTACAGAGGGGTGAGAGCCAGC 472  
 Qy 178 SerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyr 197  
 Db 473 AGTCT----- 478  
 Qy 198 IleSerLySarGlySarGlyAspGluTrpLeuAlaArgTrpLySarGlyAlaGluLys 217  
 Db 479 ---AGTCCCGCAAAAGACAGCGGACTG----- 505  
 Qy 218 AlaLyValIleAlaGlyMetAsnAlaValGluGlnGlnGlyProGlyGluSerGln 237  
 Db 506 -----CTGACAGATGCTTGGACATCTGTGGCGCCCTCAGGTCGGGGAGAGCCAA 559  
 Qy 238 LySValGluGluAlaSerProProAlaValGln-----Gln 249  
 Db 560 -----CCAGGCTCTCCGCGCTCGAGGTCACACACTCTGTGCTGTGGC 607  
 Qy 250 ProThrAspProAlaSerProThrValAlaThrThrProGluProValGlySerAspAla 269  
 Db 608 CCCCAACTTCCCGAAGATCCACCGGCTGCCCGCCACCCAGCGGGTGTGTCCCGCTC 667  
 Qy 270 GlyAspLySarAlaThrLyAlaGlyAspAspGluProGluTyrGluAspGlyArgGly 289  
 Db 668 ATGAGCGCGGTCCGGGTGCAAGGTGGAGAGCTCC----- 703  
 Qy 290 PheGlyIleGlyGluLeuValTrpGlyLysLeu-----ArgGlyPheSerTrpPro 307  
 Db 704 -----GGGAGCGGAGCGTCCCATTAAGTCTGCCCGGGGCTCTCA-----CCA 748  
 Qy 308 GlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGluGlyThrArgTrp 327  
 Db 749 GCCCGG----- 754  
 Qy 328 ValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLysLeuMetProLeu 347  
 Db 755 -----CAAGCTCTCTCCCGCC 772  
 Qy 348 SerSerPheCysSerAlaPheIleGlnAlaThrTyrAsnLysGlnProMetTyrArgLys 367  
 Db 773 TCT-----GAGAGCCCTCACTGCTCCGGGGCCCAAGTGAAGCCCTCTCCGAGGCC 823  
 Qy 368 AlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPheProValCys 387

Db 824 GCTCGGTGAGGTGAGAG----- 844  
 Qy 388 HisAspSerAspGluSerAspThrAlaValAlaGluValGlnAsnLysProMetIle 407  
 Db 845 GAGGATGCTCTAGCTCCGAGAGTCTGGGGTCCGTTCTGAAGGCCAACTCGG--- 901  
 Qy 408 GluTrpAlaLeuGlyGlyPheGlnProSerGly-----ProLysGly--- 421  
 Db 902 -----GCTCTGGTGGCGCGCGCTGAGAGAGACCCGCGTGTCCCGCGGGGCG 955  
 Qy 422 -----LeuGluProProGluGluGlnLysAsnProTyrLysGluVal 435  
 Db 956 GCAGCAGAGCGCTGCCCTGTGTCCTCAAGAAAGATTCCGCTTTCAGGCCAGGCTC 1015  
 Qy 436 TyrThrAspMetTrpValGluProGluAlaAlaAlaTyrAlaProProProAlaLys 455  
 Db 1016 -----GCCCTGTGAGACAGAGAC-----GCCCGATGGGGCGCGGGCGC 1054  
 Qy 456 LysProArgLysSerThrAlaGluLys-----ProLysValLys 468  
 Db 1055 TCCCGCTGACCAACAGCGATGATGATTTCACAGCTGCTATCTCGCTCAATCAC 1114  
 Qy 469 GluIleIleAspGluArgThrArgGluArgLeuValTyrGlu--ValArgGlnLysCys 488  
 Db 1115 GCCTTATGGCAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1174  
 Qy 488 rGAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThrLeuGlnHis-Pro 507  
 Db 1175 GGGGCTCCGAGCGCTTTC-----CCGCGCGGAGATTCACTGCTGCTGCTCAACCCG 1231  
 Qy 508 LeuPheValGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGluCysAlaTyrGln 527  
 Db 1232 GTGCTGTAGGCGAC-----TTCCCGAGCTGGGCTACCG 1267  
 Qy 528 TyrAsp-----AspAspGlyTyrGlnSerTyrCysThrIle----- 539  
 Db 1268 CCGAGCCCGCAGCCAGCAGAGACGCTACCTCTATAGCGACTTCAGCCGCCGCT 1327  
 Qy 540 -----CysCysGlyGlyArgGluVal 546  
 Db 1328 CTAAAGTAAAG 1387  
 Qy 547 LeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysValAspLeuLeu 566  
 Db 1388 CTGTGTGCGCGGTGCCAAC----- 1405  
 Qy 567 ValGlyProGlyAlaAlaGlnAlaAlaIleLysGluAspPro-TTrpAsnCysTyrMetCys 586  
 Db 1406 -----CCGCGAGCTTCCCGGATTTCCGTTGG----- 1433  
 Qy 586 sGlyHisLysGlyThrTyrGlyLeuLeuArgArgArgGluAspTrpProSerArgLeuG 606  
 Db 1434 -GGCAGCCGCGCGCGC-----TGCCTGGAGAGAGAGAGAGAGAGAGAGAGAG 1472  
 Qy 606 mMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLys-----Va 622  
 Db 1473 -----GACCCGGGAGAGAGGGGGGTGACGGCC 1498  
 Qy 622 lTyrProProValProAlaGluLysArgLysProIleArg 635  
 Db 1499 GCACCGCCAGTGCCTCAGTCTGCTGCGTCTCCGCG 1538  
 RESULT 8  
 US-09-085-1998-6  
 ; Sequence 6, Application US/090851998  
 ; Patent No. 6235879  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hayden, Michael R.  
 ; APPLICANT: Hackam, Abigail  
 ; APPLICANT: Hug, A.H.M. Mahubul  
 ; APPLICANT: Chopra, Vikramjit Singh  
 ; APPLICANT: Kalichman, Michael  
 ; TITLE OF INVENTION: Apoptosis Modulators That Interact with the

TITLE OF INVENTION: Huntington's Disease Gene  
 NUMBER OF SEQUENCES: 44  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Oppedahl & Larson  
 STREET: PO Box 5270  
 CITY: Frisco  
 STATE: CO  
 COUNTRY: USA  
 ZIP: 80443-5270  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage  
 COMPUTER: IBM compatible  
 OPERATING SYSTEM: MS DOS 5.0  
 SOFTWARE: WordPerfect  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/085,199B  
 FILING DATE:  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Larson, Marina T.  
 REGISTRATION NUMBER: 32038  
 REFERENCE/DOCKET NUMBER: UBC.P-013US2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (970) 668-2050  
 TELEFAX: (970) 668-2052  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3251  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHETICAL: no  
 ANTI-SENSE: no  
 ORIGINAL SOURCE:  
 ORGANISM: human  
 FEATURE:  
 OTHER INFORMATION: cDNA for Huntington-interacting protein

US-09-085-199B-6

Alignment Scores:  
 Pred. No.: 0.0051 Length: 3251  
 Score: 166.00 Matches: 122  
 Percent Similarity: 31.74% Conservative: 64  
 Best Local Similarity: 20.82% Mismatches: 228  
 Query Match: 3.36% Indels: 173  
 DB: 4 Gaps: 23

US-09-720-086-7 (1-912) x US-09-085-199B-6 (1-3251)

Qy 118 GlyGluGly-----AlaAlaGluThrLeuProGluAlaSerArgAla 131  
 Db 634 AGAGAGGAGTCTCAGGAGCAGCGGGTCTCGGAGAGGCTGTGGCAGCAGTTCGACGT 693  
 Qy 132 ValGluAenGlyCybCybThrProLysGluGlyArgGlyAlaProAla----- 147  
 Db 694 GTTGGCGGGCGCTGCTGC-----CGAGCCCGCGGGCATCTTGAGATGCCGTGAG 744  
 Qy 148 GluAlaGlyLysGluGlnLysGluThrAsnIleGluSerMetLysMetGluGlySerArg 167  
 Db 745 CAGCTGGACACACCCCTCGACCTGGCTGTACAGCTCCCGACAGCTACCTGGTGGAGCAG 804  
 Qy 168 GlyArgLeuArgGlyGlyLeuGlyTrpGluSerSerLeuArgGln----- 182  
 Db 805 GGC-----CCAGGAGGCGCTTGA-----TGCCGTGAGCACCCTTGAGGAGGG 846  
 Qy 183 ---ArgProMetProArgLeuThrPheGlnAlaGlyAspProTyrTyrIleSerLysArg 201  
 Db 847 CCACGCCAGTACCTGACCTCCT-TGGCAGACGCGCTCGCCCTGTGGCAGCTCTGACCC 905  
 Qy 202 LysArgAspGluTrpLeu-----Ala-ArgTrpLys----- 211  
 Db 906 GCTTCTCCACCTGGCTGCGGATACCATCATCAATGGCGGTGCCACCTCGCACCCTGGCTC 965  
 Qy 212 -----ArgGluAlaGluLysLysAlaLysValIleAlaGlyMetAsnAlaVa 227  
 Db 966 CCACGACCTGCGGACCGCTCATAGACACCTGAGGAGTGGGGCGCGGCTCTGG 1025  
 Qy 227 LgluGluAenGlnGlyProGlyLysGlnLysValGluGluAlaSerProAlaVa 247  
 Db 1026 AGCTCATGGGCGAGCTCGAGGACCGAGGCTCTCGGCACATCGACGCGAGCTGGTGC 1085  
 Qy 247 LglGlnPro-----ThrAspProAlaSerProThrVa 258  
 Db 1086 GGACACCCCTCGAGGCGATCTTTCAGCTGGGCCAAGAACTGAAACCCCAAGAGCTAGATG 1145  
 Qy 258 lAlaThr----- 260  
 Db 1146 TCGCGCAGGAGGAGCTGGGGCGCTGTCGACAAAGGAGATGGCGGCCACATCCGACGCCA 1205  
 Qy 261 -----ThrProGluProValGI 266  
 Db 1206 TTGAAGATGCTGTGCGGAGATTGAGGACATGATGAACACGAGCAGCCAGCTCGG 1265  
 Qy 266 ySerAspAlaGly-----AspLysAsnAl 274  
 Db 1266 GGGTGAAGCTGGAGGTGAACGAGAGATCTCAACTCTGCGACAGACCTGATGAAGCTA 1325  
 Qy 274 aThrLysAlaGlyAspAsp-----GluProGlu-TyrGluAspGlyArgG 289  
 Db 1326 TCGGCTCTGTCGACACATCCACTAGCTCGAGAGGATCTGTGAGAGCGGAGG 1385  
 Qy 289 lPheGlyIleGlyGluLeuValTrpGlyLysGlyLysGlyPheSerTrpTrpProGlyA 309  
 Db 1386 GGGCAGCCACGACGAGGAATTTTACGCCAAG-----AACTCGCGTGGACGAAAGGCC 1439  
 Qy 309 rgile-----ValSerTrpTrpMetThrGlyArgSerArgAlaAlaGluG 324  
 Db 1440 TCATCTCGGCTCCAAAGCTGTGGCTGGGAGCCACACAGCTGCTGGAGGAGCTGAC- 1498  
 Qy 324 lYThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLysL 344  
 Db 1499 -----AAGTGTGCTCTCACCGGCAAGTATGAGGAGCTCATCTGCTCCACAGCA 1553  
 Qy 344 euMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnProm 364  
 Db 1554 TCGAGCCAGCAGCGCCAGCTGTGGCGGCTCCCAAGGTCGAAGGCCAACACAGCAGCC 1613  
 Qy 364 erTyrArgLysAlaIleTyrGluValLeuGln-----v 375  
 Db 1614 CCCACTGAGCGCGCTCGAGGAATGTTCTCTGCACAGTCAATGAGAGGGCTGCCAATGTGG 1673

Qy 8 GlyProGlyAspThrSerSerAlaAlaGluArgGluAspArgLysAspGlyGlu 27  
 Db 313 GGAGCAGGTGAACGGGAGTCGGAGTTGAAGCTAGAGAGAGAGCGACCGAGAGAA 372  
 Qy 28 GluGlnGluGluProArgGlyLysGluGlnGluProSerThrThrAlaArgLys 47  
 Db 373 GCTCAAGAGGAGCTGGAGGCCAAGCGCGGAGAGCTGGGCCG-----CGC 417  
 Qy 48 ValGlyArgProGlyArgLysArgLys-----His 57  
 Db 418 GCAGGAGGCGCTGAGCCACACAGAGCAGAGCAAGTCGGAGCTCAGCTCAGCGCTGGACAC 477  
 Qy 58 ProProValGluSerGlyAspThrProLysAspProAlaValIleSerLysSerProSer 77  
 Db 478 ACTGAGTGGGAGAGGATGCTCTGAGTGGAGCTGTGCGCAGCGGAGGAGCAGCTGCT 537  
 Qy 78 MetAlaGluAspSerGlyAlaSerGluLeuLeuProAsnGlyAspLeuGluLysArgSer 97  
 Db 538 GCGCGGCAGAGCTGGTGGCGAGAC-----AGAGGCGGCGCT 576  
 Qy 98 GluProGlnProGluGluGlySerProAlaGlyGlnLysGlyGlyAlaProAlaGlu 117  
 Db 577 GAGCGGGAGCAGCGCCAGCTCCAGGAGCAGGGGAGTTCAGGG---CCGGCTGGC 633







[illegible]

Qy	532	yTyrGlnSerTyrCysThrIleCysCysGlyCylArgGluValLeuMetCysGlyAsnAs	552
Dd	2024	GATGGATGTGTGGCTGTGTGGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTCA	1965
Qy	552	nAasnCysCysArgCysPheCysValgluCysValaspLeuLeuValgIryProGlyAlaAal	572
Dd	1964	TGTGTGGTGTCATCGT	1916
Qy	572	aGlnAlalaIleLysGluasProTrpAsnCysTyrrMetCysGlyHis	588
Dd	1915	-----GTGTTCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1889
RESULT 11			
US-08-949-386-23			
Sequence 23, Application US/08949386			
Patent No. 6090623			
GENERAL INFORMATION:			
APPLICANT: Harpold, Michael			
APPLICANT: Ellis, Steven			
APPLICANT: Williams, Mark			
APPLICANT: McCue, Ann			
APPLICANT: Gillespie, Alison			
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND			
TITLE OF INVENTION: METHODS			
NUMBER OF SEQUENCES: 38			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Brown, Martin, Haller & McClain			
STREET: 1660 Union Street			
CITY: San Diego			
STATE: California			
COUNTRY: US			
ZIP: 92101			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: Patent In Release #1.0, Version #1.25			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/949,386			
FILING DATE:			
CLASSIFICATION:			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US/08/290,012			
FILING DATE: 11-AUG-1994			
APPLICATION NUMBER: 08/149,097			
FILING DATE: 5-NOV-1993			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: 08/105,536			
FILING DATE: 11-AUG-1993			
ATTORNEY/AGENT INFORMATION:			
NAME: Seidman, Stephanie L.			
REGISTRATION NUMBER: 33,779			
REFERENCE/DOCKET NUMBER: 519808			
TELECOMMUNICATION INFORMATION:			
TELEPHONE: (619) 238-0999			
TELEFAX: (619) 238-0062			
INFORMATION FOR SEQ ID NO: 23:			
SEQUENCE CHARACTERISTICS:			
LENGTH: 7791 base pairs			
TYPE: nucleic acid			
STRANDEDNESS: double			
TOPOLOGY: linear			
MOLECULE TYPE: DNA (genomic)			
FEATURE:			
NAME/KEY: CDS			
LOCATION: 237..7037			
OTHER INFORMATION: /standard_name= "Alpha-1A-2"			
US-08-949-386-23			
Alignment Scores:			
Pred. No.: 0.0401 Length: 7791			
Score: 161.00 Matches: 165			
Percent Similarity: 29.60% Conservative: 73			

Best Local Similarity: 20.52% Mismatches: 234  
Query Match: 3.26% Index: 334  
DB: 3 Gaps: 41  
US-09-720-086-7 (1-912) x US-08-949-386-23 (1-7791)  
Qy 7 SerGlyProGlyAspThr-----SerSerAlaAlaGluArgGlu 21  
Db 2636 AATGACCCGACGACGCTGTAAGCTGCTACACGCGACCTGCGCCGACATGAA 2695  
Qy 22 Asp-----ArgGlyAspGlyGluGluGluGluProArgGlyGlyGlu 37  
Db 2696 GACCACTGACGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2752  
Qy 38 ArgGluGluProSerThrThrAlaArgLys-----ValGlyArgProGlyArgLysArg 55  
Db 2753 CAACAGAGCG 2812  
Qy 56 LysHisProProValGluSerGlyAspThrProLysAspProAlaValIleSerLysSer 75  
Db 2813 GGACTTCTCTAG---GAAACAGCGCGCTACACGATCG-----GGCCCGGGA 2857  
Qy 76 ProSerMetAlaGlnAspSerGlyAlaSerGluLeu-----ProAsnGlyAsp 92  
Db 2858 CCCAGCGCGCTGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2917  
Qy 93 LeuGluLysArgSerGluProGluProGluGluGluSerProAlaGlyGlyGluGly 112  
Db 2918 GCTAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2968  
Qy 113 GlyAlaProAlaGluGlyGlyGlyAlaAlaGluThrLeuProGluLysArgAlaVal 132  
Db 2969 GGGGAGCT-----GGAACACCGCGGTTCTGGAGGCGAGCGCGAGCG--- 3013  
Qy 133 GluAsnGlyCysCysThrProLysGluGlyArgGlyAlaProAlaGlu----- 148  
Db 3014 -----AGCCAGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3052  
Qy 149 AlaGlyLysGluGluGluGluGluThrAsn----- 157  
Db 3053 GCGAGGCG 3112  
Qy 158 IleGluSerMetLysMetGluGlySerArgGlyArgLeuArgGlyGlyTyrGlu 177  
Db 3113 TCAGCTCATGCG 3172  
Qy 178 SerSerLeuArgGluArgProMetProArgLeuThrPheGlnAlaGlyAspProTyr 197  
Db 3173 GAGGCG 3205  
Qy 198 IleSerLysArgLysArgAspGluTyrPheAlaArgTyrLysArgGluGluLys 217  
Db 3206 CCGGCG 3259  
Qy 218 AlaLysVal-----IleAlaGlyMetAsnAlaValGluGluAsnGly 232  
Db 3260 GCACCGGATGGCGCTCCAGCACCTACGAGGAGGAGGAGGAGGAGGAGGAGG 3316  
Qy 233 ProGlyGluSerGluLysValGluGluLysSerProProAlaValGlnGluProThr 252  
Db 3317 -----GCGAGGCGATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3361  
Qy 253 ProAlaSerProThrValAlaThrThrProGluProValGlySerAspAlaGlyAsp 272  
Db 3362 -----TGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3388  
Qy 273 AsnAlaThrLysAlaGlyAspAspGluProGluTyrGluAspGlyArgGlyPheGly 292  
Db 3389 CCGGCGCATTCAGCGAGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3448  
Qy 293 GlyGluLeuValTyrGlyLysLeuArgGlyPheSerTyr-----TyrPro 307  
Db 3449 CATGAAGAAACAAGCTGGCGACCGCGGAGTGGCGCGCTCCCGACGCGAGCTTGG 3508

Qy 308 GlyArgIleValSerTyrTyrMetThrGlyArgSerArgAlaAlaGluGlyThrArg 327  
Db 3509 -----GCGCGCGCTGCCCGAGCGCGCGCGCGCGCGCGCGCGCGCG 3535  
Qy 328 ValMetTyrPheGlyAspGlyLysPheSerValValCysValGluLysLeuMetPro 347  
Db 3536 -----GATGGGAAA----- 3544  
Qy 348 SerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGluProMetTyrArg 367  
Db 3545 -----CAGCACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3552  
Qy 368 AlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPheProVal 387  
Db 3563 -----CATCTGGCGAT-----CCCTGC 3580  
Qy 388 HisAspSerAspGluSerAspThrAlaLysAlaValGluValGlnAsnLysProMet 407  
Db 3581 CATGGCCACCAACCCCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3625  
Qy 408 GluTyrAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGluProProGlu 427  
Db 3626 -----CAACCGGCGGACCATTCATCCGCGCGCGCGCGCGCGCGCG 3664  
Qy 428 GluLysAsn-----ProTyrLys-----GluValTyrThrAspMetTyr----- 440  
Db 3665 CCCCGAGATATAGCTTATTCGACCAACCCAGCGGAGGAGGAGGAGGAGGAGG 3724  
Qy 441 -----ValGluProGluAlaAlaAlaTyrAlaPro-----ProProAlaLys 455  
Db 3725 TGCCAGAGAAACCCGACCAACCAACGAGTGCATCCCGCGCGCGCGCGCG 3784  
Qy 456 LysProArgLysSerThrAlaGluLysProLysValLysGluIleLeuAspGluArg 475  
Db 3785 CCACACCGTGTACAGTGAACCAAAAGCGCAACCCAGCCACTGCCAAAAGAGGA 3844  
Qy 476 ArgGluArgLeuValTyrGluValArgGluLysCysArgAsnIleGluAspIleCys 495  
Db 3845 AGAGAAGAA----- 3853  
Qy 496 SerCysGlySerLeuAsnValThrLeuGluHisProLeuPheValGlyGlyMetCys 515  
Db 3854 -----GGAGGA----- 3859  
Qy 516 AsnCysLysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAspArgLysTyrGln 535  
Db 3859 ----- 3859  
Qy 536 TyrCysThrIleCysCysGlyGlyArgGluValLeuMetCysGlyAsnAsnAsnCys 555  
Db 3860 -----GAGGAGAGAGCA----- 3871  
Qy 556 ArgCysPheCysValGluCysValAspLeuLeuValGlyProGlyAlaAlaGlnAla 575  
Db 3871 ----- 3871  
Qy 576 IleuGluLysAspProTyrAsnCysTyrMetCysGlyHisLysGlyThrTyrGlyLeu 595  
Db 3872 -----CCGTGGGGA----- 3880  
Qy 596 ArgArgArgGluAspTyrProSerArgLeuGluMetPhePheAlaAsnAsnHisAsp 615  
Db 3881 -----AGACGCGCCCTAAGCAATGCTCCCTATAGCTCATGTTCTATCTGT 3937  
Qy 616 GluPheAspProProLysValTyrProProValProAla----- 628  
Db 3938 CCCCTTCGCGCGCTGGGCGCATTCATCTGCACTGCGGATCTTGAGATGCGATCT 3997  
Qy 629 -----GluLysArgLysProIleArg----- 635  
Db 3998 CATGTCATTCGATGAGCAGCATGCGCTGCGCGCGAGGAGCCCTGTGACGCCAAGC 4057

Qy 636 -----ValLeuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeuLy 651  
Db 4058 ACCTCGGAACAACGCTGCGATACCTTTGACTACGCTTTTACAGCGCTCTCCACTTT-- 4115  
Qy 651 sAspLeuGlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIleTh 671  
Db 4116 -----GAGATGGTGATCAAGATGATTGA 4138  
Qy 671 rValGlyMetValArgHisGlnGlyLysIleMetTyrValGlyAspValArgSerValTh 691  
Db 4139 CTGGGGCTGCTGCGATCAGGCT-----GCCTACTCCGCTGACCTCGGAATATCT 4192  
Qy 691 rGlnYHisIleGlnGluTyrGlyProPheAspLeuValIleGlyGlySerProCysAs 711  
Db 4193 -CGACTTCATGCTGCTCAGTGGGCCCT-----GGTAGCCTTTGCCCT 4233  
Qy 711 nAspLeuSerIle-----ValAsnProAlaArg---LysGlu 722  
Db 4234 -TCACTGGCAATAGCAAGAAAGACATCAACACGATTAAATCCCTCCGAGTCTCCCGG 4292  
Qy 722 YLeuTyrGlu 725  
Db 4293 GTGCTACGAC 4302

RESULT 12  
US-08-450-562-23  
; Sequence 23, Application US/08450562  
; Patent No. 6096514  
; GENERAL INFORMATION:  
; APPLICANT: Harpold, Michael  
; APPLICANT: Ellis, Steven  
; APPLICANT: Williams, Mark  
; APPLICANT: McCue, Ann  
; APPLICANT: Gillespie, Alison  
; APPLICANT: Feldman, Daniel  
; APPLICANT: Brenner, Robert  
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND  
; NUMBER OF SEQUENCES: 38  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; CITY: San Diego  
; STATE: California  
; COUNTRY: US  
; ZIP: 92101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/450,562  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/404,950  
; FILING DATE: 13-MAR-1995  
; APPLICATION NUMBER: 08/336,257  
; FILING DATE: 7-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/314,083  
; FILING DATE: 28-SEPT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/311,363  
; FILING DATE: 23-SEPT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/290,012  
; FILING DATE: 11-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/223,305  
; FILING DATE: 4-APR-1994  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/193,078  
; FILING DATE: 07-FEB-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/149,097  
; FILING DATE: 5-NOV-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/105,536  
; FILING DATE: 11-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/914,231  
; FILING DATE: 13-JULY-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/868,354  
; FILING DATE: 10-APR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/06903  
; FILING DATE: 14-AUG-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/745,206  
; FILING DATE: 15-AUG-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/620,250  
; FILING DATE: 30-NOV-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/603,751  
; FILING DATE: 08-NOV-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/482,384  
; FILING DATE: 02-FEB-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US89/01408  
; FILING DATE: 04-APR-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/176,899  
; FILING DATE: 04-APR-1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie L.  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 6362-519812  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 238-0099  
; TELEFAX: (619) 238-0062  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7791 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 237..7037  
; OTHER INFORMATION: /standard\_name= "Alpha-1A-2"  
US-08-450-562-23

Alignment Scores:  
Pred. No.: 0.0401 Length: 7791  
Score: 161.00 Matches: 165  
Percent Similarity: 29.60% Conservative: 73  
Best Local Similarity: 20.52% Mismatches: 234  
Query Match: 3.26% Indels: 334  
DB: 3 Gaps: 41

US-09-720-086-7 (1-912) x US-08-450-562-23 (1-7791)  
Qy 7 SerGlyProGlyAspThr-----SerSerAlaAlaGluArgGluGlu 21  
Db 2636 AATGACCCGACGACGCGCTGGAAGCTGCTACACGCGGCACCTGCGCCAGACATGAA 2695  
Qy 22 Asp-----ArgLysAspGlyGluGlnGluGlnGluGlnGluGlu 37  
Db 2696 CACGCACTTGACCGCGCGCTGGTGGACCCCGAGGAGACCGCAA---CAACAACAC 2752





Db 4193 -CGACTTCATAGTGTGTCAGTGGGCCCT-----GGTAGCCTTTGCCT 4233

Qy 711 nAspLeuSerIle-----ValAsnProAlaAArg---LysGI 722

Db 4234 -TCATGGCAATAGCAAGAAAGAACATCAACACGATTAATCCCTCCGAGTCTCCGG 4292

Qy 722 yLeuTyrglu 725

Db 4293 GTGCTAGGAC 4302

#### RESULT 13

US-08-984-709A-23

; Sequence 23, Application US/08984709A

; Patent No. 6320032

; GENERAL INFORMATION:

; APPLICANT: Williams, Mark E.

; APPLICANT: Staufferman, Kenneth A.

; APPLICANT: Harpold, Michael M.

; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND

; TITLE OF INVENTION: METHODS

; NUMBER OF SEQUENCES: 52

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Heller Ehrman White & McAuliffe

; STREET: 4250 Executive Square, Suite 700

; CITY: La Jolla

; STATE: California

; COUNTRY: US

; ZIP: 92037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ Version 1.5

; CURRENT APPLICATION DATA: US/08/984,709A

; FILING DATE: 02-DEC-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Seidman, Stephanie L.

; REGISTRATION NUMBER: 33,779

; REFERENCE/DOCKET NUMBER: 24735-9815 (formerly 6362-9815)

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 450-8400

; TELEFAX: (619) 587-5360

; INFORMATION FOR SEQ ID NO: 23:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 7791 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 237..7037

; OTHER INFORMATION: /standard\_name="Alpha-1A-2"

US-08-984-709A-23

Alignment Scores:

Pred. No.: 0.0401 Length: 7791

Score: 161.00 Matches: 165

Percent Similarity: 29.60% Conservative: 73

Best Local Similarity: 20.52% Mismatches: 234

Query Match: 3.26% Indels: 334

DB: 4 Gaps: 41

US-09-720-086-7 (1-912) x US-08-984-709A-23 (1-7791)

Qy 7 SerGlyProGlyAspThr-----SerSerAlaAlaGluArgGluGlu 21

Db 2636 AATGGACCGGACGAGCGCTGGAAGGTGCTACACGGCGCACCTGGCGGCACATGAA 2695

Qy 22 Asp-----ArgLysAspGlyGluGluGluGluGluGluGlu 37

Db 2696 GACGCACCTTGACCGCGCGCTGTGTGATCCCGCAGGAGAACCGCAA---CAACAACAC 2752

Qy 38 ArgGlnGluProSerThrThrAlaArgLys-----ValGlyArgProGlyArgLysArg 55

Db 2753 CAACAAGAGCG 2812

Qy 56 LysHisProProValGluSerGlyAspThrProLysAspProAlaValIleSerLysSer 75

Db 2813 GGACTTCTCTCAG---GAAACAGCGCGCTACACGATCG-----GGCCCGGGA 2857

Qy 76 ProSerMetAlaGlnAspSerGlyAlaSerGluLeu-----ProLenglyAsp 92

Db 2858 CCCAGCGGCTCG 2917

Qy 93 LeuGluLysArgSerGluProGlnProGluGluGluGluGluGluGluGluGluGlu 112

Db 2918 GCTAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2968

Qy 113 GlyAlaProAlaGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 132

Db 2969 GGGCAGGCT-----GGAGCAACCGCGGTTCTGGGAGGGCGAGCGCGCGCGCG 3013

Qy 133 GluAsnGlyCysThrProLysGluGluGluGluGluGluGluGluGluGluGluGlu 148

Db 3014 -----AGCAAGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3052

Qy 149 AlaGlyLysGluGlnLysGluThrAsn-----GGCGAGCGCGCGCGCGCGCGCG 157

Db 3053 GCAGGGGCGCAGCGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3112

Qy 158 IleGluSerMetLysMetGluGlySerArgGlyArgLeuArgGlyLeuGlyLeuGly 177

Db 3113 TCGAGCTCATCG 3172

Qy 178 SerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyr 197

Db 3173 GAGGCG 3205

Qy 198 IleSerLysArgLysArgAspGluTrpLeuAlaArgTrpLysArgGluAlaGluLys 217

Db 3206 CCGGCG 3259

Qy 218 AlaLysVal-----IleAlaGlyMetAsnAlaValGluGluAsnGlnGly 232

Db 3260 GCACCGCGCATGGCGCTCCAGCCACGTACGAGGGGAGCGCGCGCGCGCGCGCGCGCG 3316

Qy 233 ProGlyGluSerGlnLysValGluGluAlaSerProProAlaValGlnGlnProThr 252

Db 3317 -----GGCGAGCGCATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3361

Qy 253 ProAlaSerProThrValAlaThrThrProGluProValGlySerAspAlaGlyAsp 272

Db 3362 -----TGTCG 3388

Qy 273 AsnAlaThrLysAlaGlyAspGluProGluTrpGluAspGlyArgGlyPheGlyIle 292

Db 3389 CCGCCCAATCCAGCAGGACCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3448

Qy 293 GlyGluLeuValTrpGlyLysLeuArgGlyPheSerTrp-----TrpPro 307

Db 3449 CATGAAGAACAAACAGCTGGCCACCGCGAGTCCGCGCGCTCCCGCGCGCGCGCGCGCG 3508

Qy 308 GlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGluGlyThrArgTrp 327

Db 3509 -----CGCGCGCGCTGCCCGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3535

Qy 328 ValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLysLeuMetProLeu 347

Db 3536 -----CATGGGAAA----- 3544

Qy 348 SerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnProMetTyrArgLys 367

Db 3545 -----CAGCACCGACCG 3562

QY	368	AlaIleTyrGluValLeuGlnValAlaSerSerArgGlaGlyLeuLeuPheProValCys	387
Db	3563	-----CATGCTGGCCAT-----CCCTGC	3580
QY	388	HisAspSerAspGluSerAspThrAlaLysAlaValGluValGlnAsnLysProMetIle	407
Db	3581	CATGGCACCAACCCCGAGAACCCCGCAGCGCGGAGCGCCAA-----	3625
QY	408	GluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGluProGluGlu	427
Db	3626	-----CAACCCGGGGAACCCATCCATCCCGGCCCCCCCAAGAC	3664
QY	428	GluLysAsn---ProTyrLys-----GluValTyrThrAspMetTyr-----	440
Db	3665	CCCCGAGAAATAGCCTTATGCTCCACCAACCCCGGCGACCCAGACCAATTCAGCTAAGC	3724
QY	441	-----ValGluProGluAlaAlaAlaAlaTyrAlaPro-----ProProAlaLys	455
Db	3725	TGGCAGGAAACCGACCAACCAACCACTGGACATCCCCCAGCGTGCACCCCCCTCAA	3784
QY	456	LysProArgLysSerThrAlaGlyLysProLysValLysGluLeileAspGluArgThr	475
Db	3785	CCACACCGTGTCAAGTAGCAAAACCAACCCAGACCCACATGCCCAAAAAGAGAGA	3844
QY	476	ArgGluArgLeuValTyrGluValArgGlnLysCysArgAsnIleGluAspIleCysIle	495
Db	3845	AGAGAACAA-----	3853
QY	496	SerCysGlySerLeuAsnValThrLeuGluHisProLeuPheValGlyMetCysGln	515
Db	3854	-----GGAGA-----	3859
QY	516	AsnCysLysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAspAspGlyTyrGlnSer	535
Db	3859	-----	3859
QY	536	TyrCysThrIleCysCysGlyGlyArgGluValLeuMetCysGlyAsnAsnAsnCysCys	555
Db	3860	-----GGAGAGAGCGA-----	3871
QY	556	ArgCysPheCysValGluCysValAspLeuLeuValGlyProGlyAlaAlaGlnAlaAla	575
Db	3871	-----	3871
QY	576	IleLysGluAspProTyrAsnCysTyrMetCysGlyHisLeuGlyThrTyrGlyLeuLeu	595
Db	3872	-----CCGTGGGA-----	3880
QY	596	ArgArgArgGluAspTyrProSerArgLeuGlnMetPhePheAlaAsnAsnHisAspGln	615
Db	3881	---AGAGCGCCCTAAGCCATAGCTCCCTATAGCTCATTCATCCCTGTCCAGACCAA	3937
QY	616	GluLysAspProProLysValTyrProProValProAla-----	628
Db	3938	CCCCCTTGGCGCTGTGCCATTACATCCGAACCGCGCTACTTGAATGTGATCCT	3997
QY	629	-----GluLysArgLysProIleArg-----	635
Db	3998	CATGTCATTGGCATGACAGCATCGCCCTGGCGCGCAGACCCCTGTACAGCCAAAGC	405
QY	636	-----ValLeuSerLeuPheAspGlyIleAlaArgGlyLeuLeuValLeuLys	651
Db	4058	ACCTCGAACAACCTGCTCGCATCTTGACTACGTTTACAGGGCGTTCACCTT--	4115
QY	651	AspLeuGlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIleThr	671
Db	4116	-----GAGATGCGATCAAGATGATTGA	4138
QY	671	rValGlyMetValArgHisGlnGlyLysIleMetTyrValGlyAspValArgSerValThr	691
Db	4139	CTTGAGGCTGTCTTGATCAAGGT-----GCTACTTCGGTACCTTGGAATATCTT	4192

```

Oy 691 rGmLVyHs1l16LgLUlPp1yProPhsApleuVall1eGlyGlyserProCyAs 711
Db 4193 -CGACTTCATAGAGGTCACAGTGGGGCCT-----GGAGCCTTTGGCT 4233
Oy 711 nApLeuSer1le-----ValAaPProAlaArg--LyGcl 722
Db 4234 -TCACCTGGCATAGCAAAAGAAAGACATCAACAGCATTAATATCCCTCCGAGTCTCCGG 4292
Oy 722 yLeuTyGlu 725
Db 4293 GTGCTACGAC 4302

RESULT 14
US-08-450--272-23
; Sequence 23, Application US/08450272
; Patent No. 6387696
; GENERAL INFORMATION:
; APPLICANT: Harpold, Michael
; APPLICANT: Ellis, Steven
; APPLICANT: Williams, Mark
; APPLICANT: McCue, Ann
; APPLICANT: Gillespie, Allison
; APPLICANT: Feldman, Daniel
; APPLICANT: Brenner, Robert
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: California
; COUNTRY: US
; ZIP: 92101

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
  Patent Release #1.0, Version #1.25
APPLICATION NUMBER: US/08/450,272
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/404,950
  FILING DATE: 13-MAR-1995
  APPLICATION NUMBER: 08/336,257
  FILING DATE: 7-NOV-1994
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 08/314,083
    FILING DATE: 28-SEPT-1994
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 08/311,363
      FILING DATE: 23-SEPT-1994
      CLASSIFICATION: 435
      PRIOR APPLICATION DATA:
        APPLICATION NUMBER: 08/290,012
        FILING DATE: 11-AUG-1994
        CLASSIFICATION: 435
        PRIOR APPLICATION DATA:
          APPLICATION NUMBER: 08/223,305
          FILING DATE: 4-APR-1994
          CLASSIFICATION: 435
          PRIOR APPLICATION DATA:
            APPLICATION NUMBER: 08/149,097
            FILING DATE: 5-NOV-1993
            PRIOR APPLICATION DATA:
              APPLICATION NUMBER: 08/105,536
              FILING DATE: 11-AUG-1993

```

```

1      RESULT 14
2      : US-08-450-272-23
3      : Sequence 23, Application US/08450272
4      : Patent No. 6387696
5      :
6      : GENERAL INFORMATION:
7      :   APPLICANT: Harpold, Michael
8      :   APPLICANT: Ellis, Steven
9      :   APPLICANT: Williams, Mark
10     :   APPLICANT:
11     :   APPLICANT: McGue, Ann
12     :   APPLICANT: Gillespie, Alison
13     :   APPLICANT: Feldman, Daniel
14     :   APPLICANT: Brenner, Robert
15     :   TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
16     :   TITLE OF INVENTION: METHODS
17     :   NUMBER OF SEQUENCES: 38
18     :
19     : CORRESPONDENCE ADDRESS:
20     :   ADDRESSEE: Brown, Martin, Haller & McClain
21     :   STREET: 1660 Union Street
22     :   CITY: San Diego
23     :   STATE: California
24     :   COUNTRY: US
25     :   ZIP: 92101
26     :
27     : COMPUTER READABLE FORM:
28     :   MEDIUM TYPE: Floppy disk
29     :   COMPUTER: IBM PC compatible
30     :   OPERATING SYSTEM: PC-DOS/MS-DOS
31     :   SOFTWARE: Patent Release #1.0, Version #1.25
32     :   CURRENT APPLICATION DATA:
33     :     APPLICATION NUMBER: US/08/450,272
34     :     FILING DATE:
35     :
36     :   CLASSIFICATION: 435
37     :     PRIOR APPLICATION DATA:
38     :       APPLICATION NUMBER: 08/404,950
39     :       FILING DATE: 13-MAR-1995
40     :       APPLICATION NUMBER: 08/336,257
41     :       FILING DATE: 7-NOV-1994
42     :     PRIOR APPLICATION DATA:
43     :       APPLICATION NUMBER: 08/314,083
44     :       FILING DATE: 28-SEPT-1994
45     :     PRIOR APPLICATION DATA:
46     :       APPLICATION NUMBER: 08/311,363
47     :       FILING DATE: 23-SEPT-1994
48     :     CLASSIFICATION: 435
49     :       PRIOR APPLICATION DATA:
50     :         APPLICATION NUMBER: 08/290,012
51     :         FILING DATE: 11-AUG-1994
52     :         CLASSIFICATION: 435
53     :       PRIOR APPLICATION DATA:
54     :         APPLICATION NUMBER: 08/223,305
55     :         FILING DATE: 4-APR-1994
56     :         CLASSIFICATION: 435
57     :       PRIOR APPLICATION DATA:
58     :         APPLICATION NUMBER: 08/193,078
59     :         FILING DATE: 07-FEB-1994
60     :         CLASSIFICATION: 435
61     :       PRIOR APPLICATION DATA:
62     :         APPLICATION NUMBER: 08/149,097
63     :         FILING DATE: 5-NOV-1993
64     :       PRIOR APPLICATION DATA:
65     :         APPLICATION NUMBER: 08/105,536
66     :         FILING DATE: 11-AUG-1993
67     :

```

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/914,231  
FILING DATE: 13-JULY-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/868,354  
FILING DATE: 10-APR-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/06903  
FILING DATE: 14-AUG-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/745,206  
FILING DATE: 15-AUG-1991  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-519812  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 238-0999  
TELEFAX: (619) 238-0062  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7791 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 237..7037  
OTHER INFORMATION: /standard\_name="Alpha-1A-2"

US-08-450-272-23

Alignment Scores:  
Pred. No.: 0.0401 Length: 7791  
Score: 161.00 Matches: 165  
Percent Similarity: 29.60% Conservative: 73  
Best Local Similarity: 20.52% Mismatches: 234  
Query Match: 3.26% Indels: 334  
DB: 4 Gaps: 41

US-09-720-086-7 (1-912) x US-08-450-272-23 (1-7791)

QY 7 SerGlyProGlyAspThr-----SerSerSerAlaAlaGluArgGluGlu 21  
Db 2636 AATGACCCCGGACGAGCGCTGGAAGGCTGCTACACGCGGCACCTCGCGGCAGACATGAA 2695  
QY 22 Asp-----ArgLysAspGlyGluGluGluGluProArgGlyLysGluGlu 37  
Db 2696 GACGCACTTGGACCGCGCTGGTGGTGGACCGCAGGAGAACCGCAA---CAACACAC 2752  
QY 38 ArgGluGluProSerThrThrAlaArgLys-----ValGlyArgProGlyArgLysArg 55  
Db 2753 CAACAAAGAGCGGGCGCGCCAGCCACCGTGGACACGCGCTCGCGCCAGCAGCGCCCGCA 2812  
QY 56 LysHisProProValGluSerGlyAspThrProLysAspProAlaValIleSerLysSer 75  
Db 2813 GGACTTCCTCAG---GAAACAGGCGCCCTACACCATCG-----GGCCCGGGA 2857  
QY 76 ProSerMetAlaGlnAspSerGlyAlaSerGluLeuLeu-----ProAsnGlyAsp 92  
Db 2858 CCCCAGCGGCTCGCGCGGCGCTGGACGACGACGAGGCGCTGGCGGGAAGCCAGGAGCCGA 2917  
QY 93 LeuGluLysArgSerGluProGlnProGluGluGlySerProAlaGlyGlyGlnLysGly 112  
Db 2918 GCTGAGCGGGAGGAGCCCTACGCGCGCGAGTCCGACACCA---CGCCCGGGA 2968  
QY 113 GlyAlaProAlaGluGlyGluGlyAlaAlaGluThrLeuProGluAlaSerArgAlaVal 132  
Db 2969 GGGCAGCCT-----GGAGCAACCCGGGTCTTGGGAGGCGCGAGCGCG--- 3013

QY 133 GluAsnGlyCysCysThrProLysGluGlyArgGlyAlaProAlaGlu----- 148  
Db 3014 -----AGGCAAGCGCGGGAGCCCGCCACCGAGGACACGTGCACCG 3052  
QY 149 AlaGlyLysGluGlnLysGluThrAsn----- 157  
Db 3053 GCAGGGGGGACGACGAGGAGCGCGAGCGGTCCCGCGACGCGGCGCGAGGAGCA 3112  
QY 158 IleGluSerMetLysMetGluGlySerArgLysArgLysGlyLeuGlyTrpGlu 177  
Db 3113 TCGAGCTCATCGCGCGACCGCAGCGCGCGGGAGGGTCCGAGGACAGCGCGAGCG 3172  
QY 178 SerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyr 197  
Db 3173 GAGGGCGGGACCGCGA-----GGCGAGCGCGCGCGCG 3205  
QY 198 IleSerLysArgLysArgAspGluTrpLeuAlaArgTrpLysArgGluAlaGlyLys 217  
Db 3206 CCGGGCGCGGAGGCGCGGAGCGCGCGCGCGCGCGA-----GCGCAGGAGAG 3259  
QY 218 AlaLysVal-----IleAlaGlyMetAsnAlaValGluGluAsnGlnGly 232  
Db 3260 GCACCGGCATGGCGCTCCAGCCACGTACGAGGGGAGCGCG---GAGGAGGACAAGGA 3316  
QY 233 ProGlyGluSerGlnLysValGluAlaSerProProAlaValGlnGlnProThrAsp 252  
Db 3317 -----GCGAGGCGATCGGAGGAGGAGAAACAGGGTCCGGGGTCCC----- 3361  
QY 253 ProAlaSerProThrValAlaThrThrProGluProValGlySerAspAlaGlyAsp 272  
Db 3362 -----TGTCGCGGCGCCCACTGTC-----AACAC 3388  
QY 273 AsnAlaThrLysAlaGlyAspGluProGluTrpGluAspGlyArgGlyPheGlyLe 292  
Db 3389 CGGCGCAATCAGCAGGACCTGGCGCGCAAGACCCACCCCTGGCAGAGATATTGACAA 3448  
QY 293 GlyGluLeuValTrpGlyLysLeuArgGlyPheSerTrp-----TrpPro 307  
Db 3449 CATGAAGAACAAACAGCTGGCCACCGCGAGTCCGCGCGCTCCCGACGCGAGCTTTGGCCA 3508  
QY 308 GlyArgIleValSerTrpMetThrGlyArgSerArgAlaAlaGluGlyThrArgTrp 327  
Db 3509 -----GCGCGGCGCTGCGCCAGAGCCAGCCAA--- 3535  
QY 328 ValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLysLeuMetProLeu 347  
Db 3536 -----GATGGGAAA----- 3544  
QY 348 SerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnProMetTyrArgLys 367  
Db 3545 -----CAGCACCGACCCCGCGCC----- 3562  
QY 368 AlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPheProValCys 387  
Db 3563 -----CATCTGGCCAT-----CCCTGC 3580  
QY 388 HisAspSerAspGluSerAspThrAlaLysAlaValGluValGlnAsnLysProMetIle 407  
Db 3581 CATGGCCACCAACCCCAAGCAGCGCGCGCGCGCGAGCCCA--- 3625  
QY 408 GluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGluProGluGlu 427  
Db 3626 -----CAACCGGGGAACCCATCCATCCCGCGCGCGCGCCCAAGAC 3664  
QY 428 GluLysAsn---ProTyrLys-----GluValTyrThrAspMetTrp----- 440  
Db 3665 CCCCAGAAATAGCTTTATCGTCACCAACCCAGCGGCGCACCCAGCAATTCAGTAAAGAC 3724  
QY 441 -----ValGluProGluAlaAlaTyrAlaPro-----ProProAlaLys 455  
Db 3725 TGCCAGGAACCCGACCAACACACAGTGGATCCCCCGCGCGCGCGCCCTCAA 3784

```

Qy 456 LysProArgLysSerThrAlaGluLysProLysValLysGluIleLeuAspGluArgThr 475
Db 3785 CCACACCCGCTGCTAGCAACAAAGCGCAACCCAGACCCACTGCGCAAAAAGAGGA 3844
Qy 476 ArgGluArgLeuValTyrGluValArgGlnLysArgAsnIleGluAspIleCysIle 495
Db 3845 AGAAGAAGAA----- 3853
Qy 496 SerCysGlySerLeuAsnValThrLeuGlnHisProLeuPheValGlyGlyMetCysGln 515
Db 3854 -----GAGAGA----- 3859
Qy 516 AsnCysLysAsnCysPheLeuGlnCysAlaTyrGlnTyrAspAspAspGlyTyrGlnSer 535
Db 3859 ----- 3859
Qy 536 TyrCysThrIleCysCysGlyGlyArgGluValLeuMetCysGlyAsnAsnAsnCysCys 555
Db 3860 -----GGAGGAAGAGCA----- 3871
Qy 556 ArgCysPheCysValGluCysValAspLeuLeuValGlyProGlyAlaAlaGlnAla 575
Db 3871 ----- 3871
Qy 576 IleLysGluAspProTyrPaspAsnCysTyrMetCysGlyHisLysGlyThrTyrGlyLeuLeu 595
Db 3872 -----CCGTGGGGA----- 3880
Qy 596 ArgArgArgGluAspTyrProSerArgLeuGlnMetPhePheAlaAsnAsnHisAspGln 615
Db 3881 ---AGACGGCCCTAAGCCCAATGCTCCCTATAGCTCCATGTTCACTCTGTCAGACGACCA 3937
Qy 616 GluPheAspProProLysValTyrProProValProAla----- 628
Db 3938 CCCCCTTCGCGCGCTGCGCCATTCATCTGAACCTGCGCTACTTGAGATGTGATCTCT 3997
Qy 629 -----GluLysArgLysPheIleArg----- 635
Db 3998 CATGTCATTCCTCAGACAGCATGCGCTGCGCGAGGAGCCCTGTCGACCCCAACGC 4057
Qy 636 -----ValLeuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeuLys 651
Db 4058 ACCTCGGAACAAGCGTGCCTGATCTTTGACTAGCTTTTTCAGCGCGCTTCACCTTT-- 4115
Qy 651 AspLeuGlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIleTh 671
Db 4116 -----GAGATGGTGAATCAAGATGATTGA 4138
Qy 671 rValGlyMetValArgHisGlnGlyLysIleMetTyrValGlyAspValArgSerValTh 691
Db 4139 CCGTGGGCTGCTGCTGATCAGGGT-----GCCCTACTTCGCTGACCTCTGGAAATATTCT 4192
Qy 691 rGlnLysHisIleGlnGluTyrGlyLysProPheAspLeuValIleGlyGlySerProCysAs 711
Db 4193 -CGACTTCATAGTGTGTCAGTGGGGGCTT-----GGTACCTTTTGCTT 4233
Qy 711 AspLeuSerIle-----ValAsnProAlaArg-----LysGln 722
Db 4234 -TCACCTGGCAATAGCAAGAAAGACATCAACAGATTAATATCCCTCCGAGTCTCTCGG 4292
Qy 722 YLeuTyrGln 725
Db 4293 GTGCTACGAC 4302

```

```

APPLICANT: Gillespie, Alison
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
METHODS
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSER: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: California
COUNTRY: US
ZIP: 92101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/949,386
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/290,012
FILING DATE: 11-AUG-1994
APPLICATION NUMBER: 08/149,097
FILING DATE: 5-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/105,536
FILING DATE: 11-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 519808
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 238-0999
TELEFAX: (619) 238-0062
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 7808 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 237..7769
OTHER INFORMATION: /standard_name="Alpha-1A-1"
US-08-949-386-22
Alignment Scores:
Pred. No.: 0.0402 Length: 7808
Score: 161.00 Matches: 165
Percent Similarity: 29.60% Conservative: 73
Best Local Similarity: 20.52% Mismatches: 234
Query Match: 3.26% Indels: 334
Gaps: 41
US-09-720-086-7 (1-912) x US-08-949-386-22 (1-7808)
Qy 7 SerGlyProGlyAspThr-----SerSerSerAlaAlaGluArgGluGln 21
Db 2636 AATGACCCGAGCGAGCGCTGAGAGCTGCTACACCGGACCTGCGCGCAGACATGAA 2695
Qy 22 Asp-----ArgLysAspGlyGluGlnGlnGluGlnGluGlnGlyGlnGln 37
Db 2696 GACGCACTTGACCGCGCGCTGCTGTGTGACCCGCAAGAGAACCGCA---CAACAAAC 2752
Qy 38 ArgGlnGluProSerThrThrAlaArgLys-----ValGlyArgProGlyArgLysArg 55
Db 2753 CAACAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2812
Qy 56 LysHisProProValGluSerGlyAspThrProLysAspProAlaValIleSerLysSer 75
Db 2813 GCACTTCTCTAG---GAAACAGCGCGCGCTACCAAGATCG-----GGCCCGGA 2857

```

QY 76 ProSerMetAlaGlnAspSerGlyAlaSerGluLeuLeu-----ProAsnGlyAsp 92  
Db 2858 CCCACGGCTCGCGGGCCCTGGACGACGACGAGGCGCTGGCGGGAAGCCAGGAGGCGCA 2917  
QY 93 LeuGluLysArgSerGluProGlnProGluGluGlySerProAlaGlyGlyGlnLysGly 112  
Db 2918 GCTGAGCCGGGAGGACCTACGCGCCGAGTCGAGCCAGCA-----CGCCCGGA 2968  
QY 113 GlyAlaProAlaGluGlyGluGlyAlaAlaGluThrLeuProGluAlaSerArgAlaVal 132  
Db 2969 GGCGACGCT-----GGAGCAACCCGGGTCTGGAGGCGGAGGCGGAGCG--- 3013  
QY 133 GluAsnGlyCysCysThrProLysGluGluGlyArgGlyAlaProAlaGlu----- 148  
Db 3014 -----AGCAAGGCCGGGACCCCCACCGAGGCGACGTGCACCG 3052  
QY 149 AlaGlyLysGluGlnLysGluThrAsn----- 157  
Db 3053 GCAGGGGGCAGCAGGAGAGCCGCGCGGGTCCCGCGCACGCGCGCGGAGCA 3112  
QY 158 IleGluSerMetLysMetGluGlySerArgGlyArgLeuArgGlyGlyLeuGlyTrpGlu 177  
Db 3113 TCACGCTCATCGCGCACCGCAGCGCCCGGAGGAGGTCCGAGGAGCAAGGCGGAGCG 3172  
QY 178 SerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyr 197  
Db 3173 GAGGGCGCGCACCGCA-----GGCAGCCGGCGCGC 3205  
QY 198 IleSerLysArgLysArgAspGluTrpLeuAlaArgTrpLysArgGluAlaGluLysLys 217  
Db 3206 CCGGGCGCGAGGCGAGGCGAGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3259  
QY 218 AlaLysVal-----IleAlaGlyMetAsnAlaValGluGluAsnGlnGly 232  
Db 3260 GCACCGCGATGCGCTCCAGCCACGTACGAGGGGGAGCGCG-----GAGGGAGGACAAGGA 3316  
QY 233 ProGlyGluSerGlnLysValGluGluAlaSerProProAlaValGlnGlnProThrAsp 252  
Db 3317 -----GCGGAGGCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3361  
QY 253 ProAlaSerProThrValAlaThrProGluProValGlySerAspAlaGlyAspLys 272  
Db 3362 -----TGTCGCGGCGCCCAACCTGTC-----AACAC 3388  
QY 273 AsnAlaThrLysAlaGlyAspAspGluProGluTrpGluAspGlyArgGlyPheGlyLe 292  
Db 3389 CCGGCCAATCCAGCAGGACCTGGGCGCGCCCAAGACCCACCCCTGGCAGAGGATATTGACAA 3448  
QY 293 GlyGluLeuValTrpGlyLysLeuArgGlyPheSerTrp-----TrpPro 307  
Db 3449 CATGAAGAAACAAGCTGGCCACCGCGGAGTCGCGCGCTCCCGCCACGCGCGCTGGCCA 3508  
QY 308 GlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGluGlyThrArgTrp 327  
Db 3509 -----CGCGCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3535  
QY 328 ValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLysLeuMetProLeu 347  
Db 3536 -----GATGGGAAA----- 3544  
QY 348 SerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnProMetTyrArgLys 367  
Db 3545 -----CAGCAGCGACCCCGCGCC----- 3562  
QY 368 AlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPheProValCys 387  
Db 3563 -----CATGCTGGCCAT-----CCCTGC 3580  
QY 388 HisAspSerAspGluSerAspThrAlaLysAlaValGluValGlnAsnLysProMetIle 407  
Db 3581 CATGGCCACCAACCCCAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3625

QY 408 GluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGluProProGluGlu 427  
Db 3626 -----CAACCGGGGAACCCATCCCAATCCCGCGCGCGCGCGCGCGCGCGCGCGCG 3664  
QY 428 GluLysAsn-----ProTyrLys-----GluValTyrThrAspMetTrp----- 440  
Db 3665 CCCCAGGAATAAGCCTTATCGTCACCAACCCAGCGGACCCAGACCAATTCACCTAAGAC 3724  
QY 441 -----ValGluProGluAlaAlaAlaTyrAlaPro-----ProProAlaLys 455  
Db 3725 TGCCAGGAAACCCGACCAACACAGTGGACATCCCGCGCGCGCGCGCGCGCGCGCGCGCG 3784  
QY 456 LysProArgLysSerThrAlaGluLysProLysValLysGluLysGluLysGluArgThr 475  
Db 3785 CCACACCGCTCGTACAGTGAACAAACCCAGCAACCCAGACCCACTGCCAAAAAAGAGGA 3844  
QY 476 ArgGluArgLeuValTyrGluValArgGlnLysCysArgAsnIleGluAspIleCysIle 495  
Db 3845 AGAGAAGAA----- 3853  
QY 496 SerCysGlySerLeuAsnValThrLeuGluHisProLeuPheValGlyGlyMetCysGln 515  
Db 3854 -----GGAGGA----- 3859  
QY 516 AsnCysLysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAspAspGlyTyrGlnSer 535  
Db 3859 ----- 3859  
QY 536 TyrCysThrIleCysCysGlyArgGluValLeuMetCysGlyAsnAsnAsnCysCys 555  
Db 3860 -----GGAGGAAGACGA----- 3871  
QY 556 ArgCysPheCysValGluCysValAspLeuValGlyProGlyAlaAlaGlnAlaAla 575  
Db 3871 ----- 3871  
QY 576 IleLysGluAspProTrpAsnCysTyrMetCysGlyHisLysGlyThrTyrGlyLeuLeu 595  
Db 3872 -----CGTGGGA----- 3880  
QY 596 ArgArgArgGluAspTrpProSerArgLeuGlnMetPhePheAlaAsnAsnHisAspGln 615  
Db 3881 ---AGACGGCCCTAAGCCCAATGCTCCCTATAGCTTCAATGTTCTATCTCTCCACGACCAA 3937  
QY 616 GluPheAspProProLysValTyrProProValProAla----- 628  
Db 3938 CCCCCTTCGCGCGCTGTGCTTACATCTCTGAACCTGCGCTACTTTGAGATGTGCATCCT 3997  
QY 629 -----GluLysArgLysProIleArg----- 635  
Db 3998 CATGGTCATTGCCATGAGCAGCATCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4057  
QY 636 -----ValLeuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeuLys 651  
Db 4058 ACCTCGGAACACCGTGTGCTGATCTTACCTACCTTTTACAGCGCTCTTACCTTT--- 4115  
QY 651 sAspLeuGlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIleTh 671  
Db 4116 -----GAGATGGTGATCAAGATGATGA 4138  
QY 671 rValGlyMetValArgHisGlnGlyLysIleMetTyrValGlyAspValArgSerValTh 691  
Db 4139 CCTGGGCTCGCTCGATCAGGCT-----CCCTACTTCCGTGACCTCTGGAATATCT 4192  
QY 691 rGlnLysHisIleGlnGluTrpGlyProPheAspLeuValIleGlyGlySerProCysAs 711  
Db 4193 -CGACTTCATAGTGTGCTGAGTGGGGCGCT-----GCTAGCTTTTGCCT 4233  
QY 711 nAspLeuSerIle-----ValAsnProAlaArg---LysG 722  
Db 4234 -TCACTGCAATAGCAAGAAAGAAAGACATCAACACGATTAATCCCTCCGAGCTCCCGG 4292  
QY 722 yLeuTyrGlu 725

Mon Nov 25 08:18:36 2002

us-09-720-086-7.rni

Page 23

Db 4293 GTGCTACGAC 4302

Search completed: November 25, 2002, 02:59:49  
Job time : 272.557 secs

**THIS PAGE BLANK (USPTO)**



AF067972 3005 bp mRNA linear PRI 12-FEB-2001  
 LOCUS Homo sapiens DNA cytosine methyltransferase 3 alpha (DNMT3A) mRNA,  
 complete cds.  
 AF067972  
 VERSION AF067972.2 GI:12746531  
 KEYWORDS Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 3005)  
 Xie,S., Wang,Z., Okano,M., Nogami,M., Li,Y., He,W.W., Okumura,K.  
 and Li,E.  
 TITLE Cloning, expression and chromosome locations of the human DNMT3  
 gene family  
 JOURNAL Gene 236 (1), 87-95 (1999)  
 PUBMED 99365304  
 REFERENCE 2 (bases 1 to 3005)  
 Xie,S., Okano,M. and Li,E.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-MAY-1998) CVRC, Mass. General Hospital, 149 13th  
 Street, Charlestown, MA 02129, USA  
 3 (bases 1 to 3005)  
 Hata,K., Shironou,H., Sasaki,H. and En,L.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-FEB-2001) CVRC, Mass. General Hospital, 149 13th  
 Street, Charlestown, MA 02129, USA  
 REMARK Sequence update by submitter  
 COMMENT On Feb 12, 2001 this sequence version replaced gi:4927369.  
 FEATURES  
 source  
 1..3005  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2p23; between D2S171 and D2S174; 48-50 cM"  
 1..3005  
 /gene="DNMT3A"  
 230..2968  
 /gene="DNMT3A"  
 /function="de novo DNA cytosine methyltransferase"  
 /codon\_start=1  
 /product="DNA cytosine methyltransferase 3 alpha"  
 /protein\_id="AAD33084.2"  
 /db\_xref="GI:12746532"  
 /translation="MPAMPSSGPGPTSSAERERDRKDGGEERERGRKEEREPST  
 ARKVRGRGRKRRKHPVSGDPRKPAVTSKPSMAODSGASELPNDLEKSESPQPE  
 EGSPAGOKGAPABGSAATLPEASRAVENGCCTPEKGAPAEAKGKEKETTES  
 MKMGSRRLKLGAGBESSLRQRPRLTFQAGDPYISKRDEMLARWRKAEKKA  
 KVIAGNAVENOQPGESQKVEASPPAVQDPDPAVSTTPEVSGSDGDKNATK  
 AGDPEVEDRGGFGLVWGLKRGFSWNGRIYSWMTGSRRAEGRVWVWNGDG  
 KFSVCEKMLPSSFCGAFQATYKQPMVRKAIYELVYASSAGKLFPVCHDSDE  
 SDTAAVEONKPMIEMALGGFOPSGPRGPEPRKPKVEYVTDMMVDEAALYAP  
 PPAKKRKSIAEKPKYKEITIDETRELYEVOKCNIDICISGSLVNTLEHPL  
 FVGMCQCKKCFLECAQYDDDGISTCTCCGSEVLMCGNNCCCFCEVCVDLL  
 VGPAAQALIKEDVSNCMCGKGTGILRRERBPSRLQMFANNHOEDPDPVVP  
 PVPAEKRPILVNLFDGIATGLVLKDLGIQVRIYSEVCSITVGMRHOGKIM  
 YVGVRSVTQGIQIEMGFPLVIGSPENDSIYNPAKGIYSAEGRFFEFYLLHD  
 ARPEKDDRPFFMLFENVVAMGVGDKRDISFLSNPYMDIAKEVSAHRRARYWGN  
 PKMNRPLASTYNDKLELOCELEHRIAKFSVRRITTRTSNISKQKDHFPVPMNEKE  
 DILWCTEMERFEGFPHVHTDVSNMSRLRQLRLDRSWSVPVIRHLFAPLKEYFACV"

BASE COUNT 674 a 846 c 978 g 507 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.67e-188 Length: 3005  
 Score: 4935.00 Matches: 912  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-720-086-7 (1-912) x AF067972 (1-3005)  
 QY 1 MetProAlaMetProSerSerGlyProGlyAspThrSerSerAlaAlaGluArgGlu 20  
 Db 230 ATGCCCGCATCCTCCACGCGCCCGGGGACACACAGCTTCTGCTCGGAGCGGAG 289  
 QY 21 GluAspArgLysAspGlyGluGluGluGluGluProArgGlyLysGluGluArgGluGlu 40  
 Db 290 GAGGACCGAAAGACGAG 349  
 QY 41 ProSerThrThrAlaArgLysAlaGlyArgProGlyArgLysArgLysHisProProVal 60  
 Db 350 CCAGACACCCACCGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 409  
 QY 61 GluSerGlyAspThrProLysAspProAlaValIleSerLysSerProSerMetAlaGln 80  
 Db 410 GAAACCGGAGACACCCCAAAGACCTCGGGGATCTCCAACTCCCATCTGAGCCCG 469  
 QY 81 AspSerGlyAlaSerGluLeuLeuProAsnGlyAspLeuGluLysArgSerGluProGln 100  
 Db 470 GACTCAGCGGCTCAGAGCTATTACCAATGAGGACTTGAGAGACGGAGTGAGCCCG 529  
 QY 101 ProGluGluGlySerProAlaGlyGlyGlyGlyGlyGlyAlaProAlaGluGlyGly 120  
 Db 530 CCAG 589  
 QY 121 AlaAlaGluThrLeuProGluAlaSerArgAlaValGluAsnGlyCysCysThrProLys 140  
 Db 590 GGAGCTGAGAGACCTCCTGTAACCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 649  
 QY 141 GluGlyArgGlyAlaProAlaGluAlaGlyGlyGlyGlyGlyGlyThrAsnIleGluSer 160  
 Db 650 GAGGCGCGGAG 709  
 QY 161 MetLysMetGluGlySerArgGlyArgLeuArgGlyGlyGlyGlyGlyGlyGlyGly 180  
 Db 710 ATGAAGAAG 769  
 QY 181 ArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyrTyrIleSerLys 200  
 Db 770 CCGTCAAGCGGCGCATCCAGAGCTACCTTCCAGCGCGGGAGACCTTACTACACGAG 829  
 QY 201 ArgLysArgAspGluThrLeuAlaArgTrpLysArgGluAlaGluLysLysAlaLysVal 220  
 Db 830 CCGAAGCGGAG 889  
 QY 221 IleAlaGlyMetAsnAlaValGluGluAsnGlnGlyProGlyGlyGlyGlyGlyGly 240  
 Db 890 ATTGCAAGAAATGAATGCTGTGGAAAGAAACCAAGGGGCGGGAGCTTCAGAAAGTG 949  
 QY 241 GluAlaSerProProAlaValAlaGlnGlnProThrAspProAlaSerProThrValAlaThr 260  
 Db 950 GAGGCAACCCCTCCCTGCTGTGACAGACCCACCTGACCCCTCCACCTGCTGCTAC 1009  
 QY 261 ThrProGluProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAsp 280  
 Db 1010 ACGCTGAGCCCGTGGGCTCCGATCTGCGGCAAGAAATGCCAACCAAGAGCATGAC 1069  
 QY 281 GluProGluTyrGluLysArgGlyArgGlyGlyIleGlyGluLeuValTyrGlyLysLeu 300  
 Db 1070 GAGCGAGAGTACGAG 1129  
 QY 301 ArgGlyPheSerTrpTrpProGlyArgIleValSerTyrPheMetThrGlyArgSerArg 320  
 Db 1130 CCGGCTCTCTCTGCTGTGCGAGCGCGCATTTGTTGCTTGTGATGACGGCGGAGCCGA 1189  
 QY 321 AlaAlaGluGlyThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValLys 340  
 Db 1190 GCAGCTGAAGCACCCTGCGGTATGTGTTCCGAGAGCGCAATTCATCGTGTGTCT 1249  
 QY 341 ValGluLysLeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrThrAsn 360  
 Db 1250 GTTGAAGAGTGAAGCGCTGAGCTGTTTTCAGTGTCTTCCACGAGCCACGATACAC 1309



[illegible]

**THIS PAGE BLANK (USPTO)**





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

## OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 24, 2002, 12:10:37 ; Search time 68.3406 Seconds

(without alignments)  
3854.741 Million cell updates/sec

Title: US-09-720-086-6

Perfect score: 4590  
Sequence: 1 MKGDSRHNLNEEGASGYEEC.....WSVPVIRHLPAFLKDYFACE 859

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
	Ygapop 10.0	Ygapext 0.5
	Fgapop 6.0	Fgapext 7.0
Delop	6.0	Delext 7.0

Searched: 44162 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=x1h  
-O=/cgm2\_1/USPTO.spool/US09720086/runat.18112002.092957.29457/app.query.fasta\_1.4252  
-DB=Issued\_Patents\_NA -OPMT=faetap -SUFFIX=ini -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blobum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09720086 @CGN 1.1.08 @runat.18112002.092957.29457 -NCPUS=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-Ygapop=10 -Ygapext=0.5 -DELop=6 -DELEXT=7

Database :

1: Issued Patents NA:  
2: /cgm2\_6/ptodata/1/ina/5A.COMB.seq:  
3: /cgm2\_6/ptodata/1/ina/5B.COMB.seq:  
4: /cgm2\_6/ptodata/1/ina/6A.COMB.seq:  
5: /cgm2\_6/ptodata/1/ina/6B.COMB.seq:  
6: /cgm2\_6/ptodata/1/ina/backfillseq1.seq:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	862	18.8	2077	4	US-09-276-531-47 Sequence 47, Appl
2	165	3.6	6328	4	US-08-913-832A-1 Sequence 1, Appl
3	165	3.6	6328	4	US-09-249-181A-1 Sequence 1, Appl
4	156	3.4	2355	4	US-08-913-159-12 Sequence 12, Appl
5	150	3.3	2376	2	US-08-760-745-4 Sequence 4, Appl
6	144.5	3.1	6718	2	US-08-962-284-1 Sequence 1, Appl
7	132.5	2.9	6755	4	US-08-931-999-4 Sequence 4, Appl
8	127.5	2.8	1820	1	US-08-173-508-7 Sequence 7, Appl
9	127.5	2.8	1821	2	US-08-265-310-7 Sequence 7, Appl
10	127.5	2.8	1821	3	US-08-951-742-7 Sequence 7, Appl
11	127	2.8	2623	2	US-08-973-675-1 Sequence 1, Appl
C	125.5	2.7	34063	4	US-09-453-702B-96 Sequence 96, Appl

13	125.5	2.7	4411529	4	US-09-103-840A-1	Sequence 1, Appl
14	125	2.7	4657	3	US-09-254-325-1	Sequence 1, Appl
15	125	2.7	5643	3	US-09-079-415-5	Sequence 5, Appl
16	125	2.7	5643	3	US-08-750-458A-1	Sequence 1, Appl
17	124	2.7	4181	4	US-09-393-569-1	Sequence 1, Appl
18	123.5	2.7	28804	2	US-08-592-874-1	Sequence 1, Appl
19	123.5	2.7	28804	3	US-09-096-942-2	Sequence 2, Appl
20	123.5	2.7	28804	3	US-09-096-867-2	Sequence 2, Appl
21	123.5	2.7	4403765	4	US-09-103-840A-2	Sequence 36, Appl
22	123	2.7	3044	4	US-09-152-060-36	Sequence 2, Appl
23	123	2.7	36519	3	US-08-973-137-2	Sequence 2, Appl
24	122	2.7	8931	3	US-09-051-019-1	Sequence 1, Appl
25	121.5	2.6	5965	4	US-09-362-36A-1	Sequence 1, Appl
26	121.5	2.6	6028	4	US-09-362-336A-3	Sequence 3, Appl
C	121.5	2.6	20235	1	US-07-642-734C-3	Sequence 3, Appl
C	121.5	2.6	20235	1	US-08-439-009A-3	Sequence 3, Appl
C	120.5	2.6	1170	3	US-09-307-623-1	Sequence 1, Appl
30	119	2.6	11471	4	US-09-504-358-16	Sequence 16, Appl
31	119	2.6	11471	4	US-09-954-314-16	Sequence 16, Appl
32	118.5	2.6	1865	4	US-09-370-253-5	Sequence 5, Appl
33	118.5	2.6	6463	2	US-08-962-284-3	Sequence 3, Appl
34	118	2.6	1970	1	US-08-028-463-14	Sequence 14, Appl
35	118	2.6	1970	1	US-08-461-836-14	Sequence 14, Appl
36	118	2.6	43280	2	US-08-804-227C-1	Sequence 1, Appl
37	117.5	2.6	11604	4	US-09-385-028-13	Sequence 13, Appl
38	117.5	2.6	15079	4	US-09-385-028-13	Sequence 1, Appl
39	117	2.5	2905	4	US-08-942-686-1	Sequence 1, Appl
C	117	2.5	4265	4	US-09-061-709-1	Sequence 1, Appl
C	116.5	2.5	43280	2	US-08-804-227C-1	Sequence 1, Appl
C	116	2.5	49377	1	US-08-764-233A-1	Sequence 1, Appl
43	115.5	2.5	1800	1	US-08-139-937-11	Sequence 11, Appl
44	115.5	2.5	1800	5	PCT-US93-11310-11	Sequence 11, Appl
C	115	2.5	50937	4	US-09-428-517-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-09-276-531-47  
Sequence 47, Application US/09276531  
Patent No. 6183368  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Yue, Henry  
APPLICANT: Reddy, Roopa  
APPLICANT: Guegler, Karl J.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276,531  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/079,677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:





```

Db 451 -----TCTGCTCAGCTCTCGGAAGACTGGGGCATGGAAACATGTGAC-----CA 494
Qy 211 nValAspAlaGluSerArgAspGlyAspSerThrGluTyrGlnAspArgLysGluPheG1 231
Db 495 CGTGTCTTCAGAGAGAGATTTCGAACCTTCAACCACTAC-----534
Qy 231 yLleGlyAspLeuValTyrGlyLysIleLysGlyPheSerTrpTrpProAlaMetValVa 251
Db 535 -----AAGGCTTCAGCCAGTTTGACACCCCTCAT 566
Qy 251 lSerTrpLys-----AlaThrSerLysArgGlnAlaMetProGlyMetAlaGTr 267
Db 567 TGCCTGCAAAAATCCCAAGATTGCTGTCTCCAAAGATGATGATGTTGGTGGCAAAATG 626
Qy 267 pValGlnTrpPheGlyAspGlyLysPhe-----SerGluIleSerAlaAs 282
Db 27 GCGGGAGTTCCAGTCCAAATACCCCTTCAAGGCACTTGGGGCATGATGGCAGCTGC 686
Qy 282 pLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsnLysLe 302
Db 687 GCGACGACGACGGGTAGCTGTGTGGAG-----AGCATGTGTACAGCCATGAGGT 737
Qy 302 uValSerTyrArgLysAlaMetTyrHisThrLeuGluLysAlaArgValAlaGlyLys 322
Db 738 TGACACCAACCTCCCTCTGGAGGTGCTCTCCGCAAGGCGCAAGCAAGAGAGGCGCA 797
Qy 322 sThrPheSerSerProGlyLysLeuSerLeuGluAspGlnLeuLysProMetLeuGluTr 342
Db 798 A-----GTTCCCAATGCTCGGAGAGAGCCCAAGGCGACCTCTGTACTCTGA 845
Qy 342 pAlaHisGlyGlyPheLysProThrGlyLleGluGlyLeuLysProAsnLysLysGlnPr 362
Db 846 TGCCAG-----AAGCT-----AAACCCAAAGAAAGTACCTC 878
Qy 362 oValValAsnLys-----SerLysValArgArgSerAspSerArgAsnLe 377
Db 879 CCTGAATATCAAGCTGGAGGTTTGTGTTCCAAAGCGTAAAGATCTCGATGAGGAT-- 936
Qy 377 uGluProArgArgArgLysAsnLysSerArgArgArgThrHisAsnAspSerAlaAlaSe 397
Db 937 -----GATGACTTAGATGTGAATCTGACTTCCAGTATGTCAGT----- 975
Qy 397 rGluSerProProProLysArgLeuLysThrAsnSerTyrGlyLysAspArgGlyL 417
Db 76 -----ATCAATAGCTATTCTGTTTGTGATGTTCCAC 1007
Qy 417 uAspGlnGluSerArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuG1 437
Db 1008 CAGCCGTAGTAGCGGACCGCAAGAACTCCGAACCACTAAAAAGAAAAGAAAGAGCGCA 1067
Qy 437 uAspArgCysLeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluG1 457
Db 1068 GGAGAGGTGACTCTGTG-----1086
Qy 457 yGlyLeuCysGlnSerCysArgAspArgPheLeuGluLeuPheTyrMetTyrAspGluAs 477
Db 1087 -----GA 1088
Qy 477 pGlyLys-----GlnSerTyrCysThrValCysCysGlnGluYarGluLeuLe 493
Db 1089 TGGTTATAGACACAGACCAACAGAGACTATTGGGAGGTGTGCCAGCAAGCGGTGAGATCAT 1148
Qy 493 uLeuCysSerSerThrSerCysArgCysPheCysValGluCysLeuGluValLeuVa 513
Db 1149 CCGTGTGTAT-----ACCTGTCCCGGTGCTTACCACTGTCTGCGTGTGAT----- 1194
Qy 513 lGlyAlaGlyThrAlaGlyuAspAlaLysLeuGlnGluLys-----ProTrpSerCysTyrMetCys 532
Db 1195 -----CCGACATGAGAGAGGCTCCGAGGAGCAAGTGGAGCTGCCACACTG 1241
Qy 532 sLeuProGlnArgCysHisGlyValLeuAsnArgArgArgLysAspTrpAsnMetArgLeuG1 552
Db -----

```

```

Db 1242 CGAGAGAGAA-----GGCATCCAGTGGAGAACTTAAAGAGACAAATTCCGAGGCTGA 1292
Qy 552 nAspPhePhe-----ThrThrAspProAspLeuGluGlu 563
Db 1293 GGAGATCTCTGGAAGAGGTTGGGGAGAGACCTCTGAAAGAG 1329

RESULT 3
US-09-249-181A-1
/ Sequence 1, Application US/09249181A
/ Patent No. 6440679
/ GENERAL INFORMATION:
/ APPLICANT: Seelig, Hans Peter
/ APPLICANT: Renz, Manfred
/ TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
/ FILE REFERENCE: 8484-0059-999
/ CURRENT APPLICATION NUMBER: US/09/249,181A
/ PRIOR APPLICATION NUMBER: US 08/913,832
/ PRIOR FILING DATE: 1998-01-12
/ PRIOR APPLICATION NUMBER: PCT/DE96/00444
/ PRIOR FILING DATE: 1996-03-08
/ NUMBER OF SEQ. ID NOS: 2
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 1
/ LENGTH: 6328
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)...(5736)
US-09-249-181A-1

Alignment Scores:
Pred. No.: 2,19e-05 Length: 6328
Score: 165.00 Matches: 122
Percent Similarity: 33.90% Conservative: 79
Best Local Similarity: 20.57% Mismatches: 191
Query Match: 3.59% Indels: 202
DB: Gaps: 28

US-09-720-086-6 (1-859) x US-09-249-181A-1 (1-6328)
Qy 9 AsnGluGluGluGlyAlaSerGlyTyrGluGluCysIleLeuValAsnGlyAsnPheSer 28
Db 43 AGTAGAGAGAGAGATGATGATGCA-----CTTTGAACAACAGCTGCGCC 87
Qy 29 AspGlnSerSerAspThrLysAspAlaProSerProProValLeuGluAlaIleCysThr 48
Db 88 CCACCCCAACCCAGAAATGAAGAGACCCAGAGAGGATTTGTCAGAA-----ACA 138
Qy 49 GluProValCysThrProGluThrArgGlyArgArgSer-----61
Db 139 GAG-----ACTCCAAAGCTCAAGAGAGAGAAAGAAAGCTTAAGAAACCTCGGACCTT 189
Qy 62 -----SerSerArgLeuSerLysArgGluValSerSerLeuAsnTyrThrGln 78
Db 190 AAATCCCTAAGACCAAGCGCCAAAGAGAGCGTATGCTTATGCGGACAGCTGGGG 249
Qy 79 AspMetThrGlyAspGly-----AspArgAspAspGluVal-----Asp 91
Db 250 GACAGCTCTGGGGAGGGGCGCAGAGTTTGTGGAGAGAGAGAAAGAGGTGGCTCTCGCTCA 309
Qy 92 AspGlyAsnGlySerAspLleLeuMetProLysLeuThrArgGluLysArgThrArg 111
Db 310 GACAGTAGGCGCACGACTATACCTCTGGCAAGAGAGAAAGAGAAAGAGCTTGGACCTAAG 369
Qy 111 gThrArgSerGluSerProAlaValArgThrArgHisSerAsnGlyThrSerSerLeuG1 131
Db 370 A-----AA 372
Qy 131 uArgGlnArgAlaSerProArgLleThrArgGlyArgGlnGlyArgHisIleValGlnG1 151
Db 373 GAGAGAGAGAGCAAT-----388

```



OTHER INFORMATION: /note="The sequence shows 60 % identity and 76 % similarity  
OTHER INFORMATION: with the Bep6i methylase."  
US-08-913-159-12

## Alignment Scores:

Pred. No.:	3.18e-05	Length:	2355
Score:	156.00	Matches:	78
Percent Similarity:	35.48%	Conservative:	43
Best Local Similarity:	22.87%	Mismatches:	124
Query Match:	3.40%	Indels:	96
DB:	4	Gaps:	16

US-09-720-086-6 (1-859) x US-08-913-159-12 (1-2355)

```
Qy 581 ILeaRgVAlleuSerLeuPheaspGlyllea1aThnrglyTyLeuValleuLeuSglueu 600
Db 1395 TTGAAAATGCTCTTTTTCGCCGAGATTGGCGAATTGATTGTTGAAAATGCA 1454
Qy 601 G1y1leuValG1uLyTyrlleAlaSerGluValCySa1aG1uSerlleAlaValG1y 620
Db 1455 GGTTCAAAACA-----ATATATGCTAATGAAATT-----GATAATTATGCTGAT 1502
Qy 621 ThrValySh1sG1uG1yG1leuTyTyryValaAspValArgLy1leThrltyls 640
Db 1503 ACTTTTGAATGAAGACTTTCAGCTTAAGTA-----GACCGACGTGATTAATGATGTA 1556
Qy 641 Asn1leg1uTPrG1yProPheaspLeuVal1leg1ySerProCySaAspLeu 660
Db 1557 CAAGTGATGAATCCAGATTTCATATATATGATGATGCAAGTTTCCTTCCAGCCTTT 1616
Qy 661 SerAsnValaenProAlaArgLySglueu-----Tyrg1uG1yThrg1yArgLeuPhe 678
Db 1617 TCT--ATTGCTGTTATTCGTCAGAGCTTAAAGATGAACAAGTCGAGTATCTTTT 1673
Qy 679 PheG1uPheTyrlleuLeuAsnTyrlhArgProLySgluLyAspAsnArgProhe 698
Db 1674 TTGAACTTGTCTGATTTTGAACAAAACAAACCTCGTTGCA----- 1718
Qy 699 PheTrpMetPheGluAsnValaVala1aMetLySValaAspLyLyAsp1leSerArg 718
Db 1719 -----TTCTTTGAAAAGTTTAAATCTGTTCTCAGATAGCGGGAACATTGCA 1772
Qy 719 PheLeuAlaCySaenProValMet1leAspAla1leLySValaSer1aAla1h1sArg1a 738
Db 1773 GTTATTGCT-----TTCGAGTTAGAAAGACTA 1799
Qy 739 ArgTyT-----PheTrpG1yAsnLeuProGlyMet 748
Db 1800 GGGTACCAAGTATCTTTTCAAGCTTTAATGCTTCTGAAATATGAAATATACCT--CAA 1856
Qy 749 AsnArgPro-----ValMetAlaSerLySAspAsp----- 758
Db 1857 AATAGAGAACGATCTATATATTTGCTTTCAAAAATTAATAAAGATTATGCAAAATTTGAA 1916
Qy 759 -----LysLeuG1uLeuG1nAspCySLeuG1uPheSerArg-- 770
Db 1917 CTACCAAAATCTATACCTTTAAAAAACAAGATTCAAGATGATTGATTTCTTAAAAA 1976
Qy 770 ----- 770
Db 1977 CAAGACGATAGTCTACTATACCTCTGAAAAGAAATTTTGTGATGATTAAGAA 2036
Qy 771 -----Thra1aLySLeuLySValaG1ThrlleThrltyls 783
Db 2037 AATATGACTAATACAGACACTACATATCAGTGGCGTAAAGTTTATGTAAGAAAAAACA 2096
Qy 784 SerAsnSer1le-----ArgG1nG1yLySAsnG1nLeuPhe 795
Db 2097 AGTAATTATGATCAACAACATAACGCTAATATGGAACAGTGGGATATAT-----GTG 2150
Qy 796 ProValaValMetAsnG1yLySAspAspValleuTrpCySThnG1uLeuG1u-----Arg 813
Db 2151 CCTATATATCTTACATATATAGCGGAGATATTCGTAATTAAACCAAGAGAAATGCTTTAAC 2210
```

```
Qy 814 ILePheG1yPheProAla1h1sTyTThraSValSerAsnMetG1yArgG1yAlaArgG1n 833
Db 2211 GTTCAGAGTTTCCCAAAAGATAT--AACTTCCAACCAAGTAATGAGATTTATAT 2267
Qy 834 LysLeuLeuG1yAsnSerTrpSerValProVal1leArg1h1sLeuPheAlaProLeuLys 853
Db 2268 AAACCAACGAGAAACAGTGTGTTAGTACCAAGTTATGAAGAAATGCAAAAAATCTTGCA 2327
Qy 854 Asp 854
Db 2328 GAT 2330

RESULT 5
US-08-760-745-4
Sequence 4, Application US/08760745
Patent No. 5972658
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goll, Suiya K.
APPLICANT: Murry, Lynn E.
TITLE OF INVENTION: LUNG GROWTH FACTOR VARIANT
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/760,745
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PR-0169 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2376 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 598956
US-08-760-745-4

Alignment Scores:
Pred. No.: 0.000126 Length: 2376
Score: 150.00 Matches: 83
Percent Similarity: 36.62% Conservative: 58
Best Local Similarity: 21.56% Mismatches: 163
Query Match: 3.27% Indels: 81
DB: 2 Gaps: 15

US-09-720-086-6 (1-859) x US-08-760-745-4 (1-2376)
Qy 142 G1yATgG1nG1yATG1h1sValG1nG1uTyTProValG1uPheProAlaThraGser 161
Db 142 G1yATgG1nG1yATG1h1sValG1nG1uTyTProValG1uPheProAlaThraGser 161
```



QY 47 CysThrGluProValCysThr-----ProGluThrArgGly 58  
Db 1502 TCATCGGAGCCCAATTGCAGGCTCTGCTCAACAAAAGCCTTAGACAGGACGACAA 1561  
QY 59 ArgArgSerSerArgSerLeuSerLeuArgGluVal--SerSerLeuLeuAsnThrArg 78  
Db 1562 ATGAGGCGAAGAGCTTGGCTGAGCAGAGAGAACTGATTCTCCACTTGAAGCTTGCGCC 1621  
QY 78 LysMetThrGlyAspGlyAspArgAspArgGluValAspAspGlyAsnGlySerAspI 98  
Db 1622 AAAAGTTGGCC----- 1632  
QY 98 LysMetCProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProA 118  
Db 1633 --TTGTTTAACAAAATTGTCCAGCCAGCTC-----TCAAAAG 1666  
QY 18 LysValArgThrArgHisSerAsnGlyThrSerSerLeuGluArgGluValAspAspProA 138  
Db 1667 CGATTTCCTACCCGAGACAAATAGACAGC-----AGACAGAGGAGA----- 1707  
QY 138 rgLThrArgGlyArgGlnGlyArgHisValGlnGluLysProValGluPheProA 158  
Db 1708 -----ATGACGCTGCTATCAACTCAG-----CCAG 1735  
QY 158 LThrArgSerArgArgArgArgLaserSerSerLaserThrProTrpSerSerProA 178  
Db 1736 TCACACTCGGAGAGAGTGGAGAGAGTGCAGAGTGAAGCTCATTCCTTCCACTGCGC 1795  
QY 178 LAspValAspPheMetGluGluValThrProLysSerValSerThr-----ProSerV 196  
Db 1796 TGAACACA-----TCAGTCTTACCGTAGACATCCACGG 1828  
QY 196 aLAspLeuSerGlnAspGlyAspGlnGluLysMetAspThrThrGluValAspLaglus 216  
Db 1829 TTGCTCCAAATGTATGCGGAGATCTTGACAAAGCCACTTGGACACACATGCAAGTG 1888  
QY 216 eArgAsp-----GlyAspSerThrGluLysArgLysAspLysGluPheGlyLeuG 233  
Db 1889 CCACTGACTTAAGATTCTTCTTCAATAGAAATTCGGACTTCCAGTTAGAAAGCATTT- 1947  
QY 233 LysAspLeuValTrpGlyLysLeuLeuGlyPheSerTrpTrpPro----- 247  
Db 1948 -----CTGAATTCGCAAGCTTGGACCTTGGTAGAGGTAGCG 1987  
QY 48 -----AlaMetValValSerTrpLysAlaThrSerLysArgGlnAlaMetProGlyM 265  
Db 1988 AGAACAAGGAGATGTGAGATATGAGACAGACAGAAAGCAGAGCTTGGACAGGTC 2047  
QY 265 eArgTrpValGlnTrpPheGlyAsp-----GlyLysPheSerGluI 279  
Db 2048 GA-----GACAGTGGGATGAGAAATATGCGTCTTTGAGGAAG 2086  
QY 279 LysSerAlaAsp----- 282  
Db 2087 CAGAAGCATCTACCCATCTGAACAGAGCGAGGAAAGAGACCATTAAGAAATCTA 2146  
QY 283 -----LysLeuValAlaLeuG 288  
Db 2147 AATATGCTGTTCCAGAGAGAAAGCTGGAACGGGCGAACCTTCCATCCATCCACCTCG 2206  
QY 288 LysLeuPheSerGlnHisPheAsnLeuAlaThrPheAsn----- 300  
Db 2207 GGGATGAACCGAAGAAATTTTCCATGGCTAAATGAATGACAAAGAACTTGAACCTTGA 2266  
QY 301 -----LysLeuValS 304  
Db 2267 GGGACAGGCTCCCTTTGAAGAGAAAGTGAAGTGAAGATGTTATGAAGAAAGATT 2326  
QY 304 eTrpArgGlyValAlaMetCys-----HisThrLeuGluLysAlaArgValArgLaglyL 322  
Db 2327 CACTAAGAGCGGAGAGATTTCGGGAGACCACTTCCGAGACAGAGCGGAGCAAGCTCTGGGA 2386

QY 322 YThrPheSerSerSer-----ProGlyGluSerLeuGluAspG 335  
Db 2387 AAATATTGCTCAAAACACAGACCCCGTGTCTGGAAGCCCGGAGATTCTTGGAGACG 2446  
QY 335 LysLeuLys-----ProMetLeuGluTrpAlaHisGlyGlyPheLys---P 349  
Db 2447 CACAGAGAAAGCTCTGCAAGATTCATGTCCGATGTTGCTGCTGAGAGATCAAAAAGC 2506  
QY 349 roThrGlyLeuGlyLeu-----LysProAsnLys-----LysGlnProV 363  
Db 2507 CGACAGGG---GAGGGCTTCTTGACTCACCAGCAAAACCATGTCTATTAAGAAAGAT 2563  
QY 363 aLysLeuLysSerLysValArgArgSerSerArgAsnLeuGluProValArgArgG 383  
Db 2564 TGGCAGCTGTGAAGAAAGCGGAGAGAAATTTGAGAAACAACTCAGCAGAGAGCGAG 2623  
QY 383 Lu---AsnLysSerArgArgArgThrThrAsnAspSerAlaAlaSerGluSerProPro 402  
Db 2624 AGGCGGCAAGGCGCGCGCGGAGAGCTGCACACCCAGAGAGAGGCGGCTCCCTCATCA 2683  
QY 402 roLysArgLeuLysThrAsnSerTyrGlyLysAspArgGlyLysAspGluGluSerA 422  
Db 2684 AGAAGCGGCTC-----ACAGAAAGTCAGAGAGGCAATAGACGA 2722  
QY 422 rgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCysLeuS 442  
Db 2723 TTGAGAGAGAGAAAGCAAGCTCATCATCTGT--GAGAGAGAGGCTCGAAGAGAGAGCACA 2781  
QY 442 eArgGlyLysLys---AsnProValSerPheHisProLeuPheGluGlyGlyLeuGlyCysG 461  
Db 2782 GGAAGCGGCAAGCACTGCAGCCAGTT-----CACTGTGCTGGCAGAGATGT----- 2828  
QY 461 LysArgCysArgAspArgPheLeuGluLysPheTyrMetCysTrpAspGluAspGly----- 478  
Db 2829 -----GAAGAAAGTTT---GGGCTACCTATCTGCCATTAACCCAGTAGCCTCAGCC 2877  
QY 479 -----TyrGlnSerT 482  
Db 2878 ATTTGCGGTAAACAAAGAGGACACACCCGTTTCCAAACCCCTGGAAGATATGAGAGCC 2937  
QY 482 YrcYThrValCysCysGlu---GlyArgGluLeuLeuLeuCysSerAsnThrSerCysC 501  
Db 2938 AGACCAATATGCAAGTTGAATCGAGCTGAACTTGGACAGCTGGAACCTTTCT----- 2993  
QY 501 YsaArgCysPheCysValGluCysLeuGluValLeuValGlyAlaGlyThrAlaGluAspA 521  
Db 2994 -----AACAGGCTGAATTAAC 3009  
QY 521 LysLeuGlnGluProTrpSerCysTyrMetCysLeuProGlnArgCysHisGly-Val 540  
Db 3010 AAGTTGGCGGAGATGCCAGAAAGGTGCTCATCTGCACCGGCAAAATGCTGAAGAGAGTG 3069  
QY 541 LeuArg-----ArgArgLysAspTrpAsnMet 549  
Db 3070 ATGAAGCCAGATGATGTAACCTTTGCCAAATTTTACCGCAGCGGATGTTAATATAG 3129  
QY 550 ArgLeuGlnAspPhePheThrThrAspProAspLeuGluLys---PheGluPro----- 566  
Db 3130 CCA---AGAAGTCTCTGAGATGAGATGAGAGCTTCGATGATTTTCATCTTATGCA 3186  
QY 567 ProLysLeuTrpProAlaLeuProAlaLeuAlaLysArg----- 578  
Db 3187 CCCAAAATTTGACGTCTTCCGCGCGGAGACACAAGCGGCACTTAGGCCCAAGCGCGGCTT 3246  
QY 579 -----ArgProLeuArgValLeuSerLeuPheAspGlyLeuAlaThrGlyTyr 594  
Db 3247 CAGGCTCCAAAAAACCCTCGAAATGCTGCGGCAAGAAAGATCTCTTCAAGAAATAC 3306  
QY 595 -----LeuValLeuLysGluLeuGlyLysValGluLysTyr 607  
Db 3307 ACTGAGCAGAGATTAAAGCTTGCTTCATGAGGTCAAAAGCGGATGAAGAAAGTGAAGATG 3366  
QY 608 IleAlaSerGluValCysAlaGluSerIleAlaValGlyThrValLysHisGluGly--- 626

```

Db 3367 TCATCAACTCACTCACTTCAGAAAGTCACTCGCGGTTAGCCAGTAAGAAAGAACTTC 3426
QY 627 -----GlnIleLysThrValAsnAspValArgLysIleThrLysLysAsnIleGlu 643
Db 3427 AGCAAGCTCAGCTCGGAGCGTCAAC-----CTGACGGAACAACTCAAC 3474
QY 644 GluTrpGly---ProPheAsp-----LeuValIleGlyGlySerProCysAsn 658
Db 3475 AACAGCGCGTCCCTACAGAGCTGATCTGCTTCCAGATTAAAGGAAGACATGTG 3534
QY 659 AspLeuSerAsnValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhe 678
Db 3535 CAGACAGCGCTGGAACTTCGAGCTTCGCGCTCAACAGTGG----- 3579
QY 679 PheGluPheTyrHisLeuLeuAsnTyrThrArgProLysGluGlyAspAsnArgProPhe 698
Db 3580 ---GACTGCTCTCCCTCTCTCT-----CCCACTGCTGCTTC 3615
QY 699 PheTrpMet-----PheGluAsnValVal-----AlaMetLysValAsnAsp 712
Db 3616 CTGTGGTAGGAGAGTTTGCAAAGCTCATAGAAAGCGGAGCGCTCAGAACTTGCAACT 3675
QY 713 LysLysAspIleSerArgPheLeuAlaCysAsnProValMetIleAspAlaIleLysVal 732
Db 3676 TTAATTTCAGACAAAGAGGAACCTTGGTTGTAGAGCTACTTATATCCAAACCATTGAAGAA 3735
QY 733 -----SerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsn 749
Db 3736 GGAATTATACACACACTCATGCAGCCAAAGACTTCTGGAAGCTTCTGGTGGCCAAACC 3795
QY 750 ArgProValMetAlaSer-----LysAsnAspLysLeuGluLeuGlnAspCysLeuGlu 767
Db 3796 AGTTACCAATCTGCTGGAGCCCAAGAAAGATGAATC----- 3834
QY 768 PheSerArgThrAlaLysLeuLysLysValGlnThrIleThrThrLysSerAsnSerIle 787
Db 3835 -----TATGAAGCAGCCATTAATAGAACTTAATCGATT 3867
QY 788 ---ArgGlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysAspAspValLeu 806
Db 3868 TACCGTCTCATGGATGACAAACTTGTTCCT-----GATGACGACTAC 3909
QY 807 Trp-----CysThrGluLeuGlu-----ArgIlePheGlyPhe 817
Db 3910 TGGGGGAAATTTCCGAAGTGTCTCCCTCTGCAACCCAAAGAGGTACTGGTGTGATTTT 3969
QY 818 ProAlaHisTyrThrAspValSerAsnMetGlyArgGlyAlaArgGlnLysLeuGly 837
Db 3970 GGTAGTGAAGTT-----TACGTATGCGATGGGAAGAGTACACTAGCACA 4017
QY 838 ArgSerTrpSerValProValIleArgHisLeuPheAlaProLeuLysAspTyrPheAla 857
Db 4018 CGAAANTAGCATTTACGTGGCAGGCACTTATGGAATGGNACCTTTGACTATGAGAAC 4077
QY 858 CysGlu 859
Db 4078 TGTGAC 4083

RESULT 7
US-08-931-999-4
; Sequence 4, Application US/08931999
; Patent No. 6043219
; GENERAL INFORMATION:
; APPLICANT: Iandolo, John J.
; APPLICANT: Crupper, Scott S.
; TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hovey, Williams, Timmons & Collins
; STREET: 2405 Grand Boulevard, Suite 400
; CITY: Kansas City
; STATE: Missouri

```

```

; COUNTRY: U.S.A.
; ZIP: 64108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,999
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/710,561
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26,262
; REFERENCE/DOCKET NUMBER: 25043-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 816/474-9050
; TELEFAX: 816/474-9057
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6755 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: UT0007
; US-08-931-999-4

```

```

Alignment Scores:
Pred. No.: 0.0394
Score: 132.50
Percent Similarity: 35.01%
Best Local Similarity: 19.90%
Query Match: 2.89%
Indels: 112
Gaps: 13

```

US-09-720-086-6 (1-859) x US-08-931-999-4 (1-6755)

```

QY 26 AsnPheSerAspGlnSerSerAspThrLysAspAlaProSerProValLeuGluAla 45
Db 4503 AACCAAGCAGGAGGACCAACACACAAAGAGCGGAG----- 4541
QY 46 IleCysThrGluProValCysThrProGluThrArgGlyArgSerSerSerArgLeu 65
Db 4542 -----ACGGAAGAAAA--ACACCCGAAAAAACACAAAGAACACACAAACCA 4592
QY 66 SerLysArgGluValSerSerLeuLeuAsnTyrThrGlnAspMetThrGlyaspGly--- 84
Db 4593 ACCCAAGAGAGAAAGA-----AACACAAAGAAAAAGAAAAACCGGAGGAAGA 4643
QY 85 -----AspArgAspAspGluValAspAspGlyAsnGlySerAspIleLeuMetPro 101
Db 4644 GGAACACAGCCAGAGAAACACACACACACGAGGAGGAGGACCA----- 4698
QY 102 LysLeuThrArgGluThrLysAspThrArgThrArgSer--GluSerProAlaValArgT 121
Db 4689 -----GAACAAAAAGAGGCAAAACACAGAGGAGGACAAACCCAGCAACAGCA 4736
QY 121 hrArgHisSerAsnGlyThrSerSerLeuGluArgGlnAlaSerProArgIleThrA 141
Db 4737 ACAAGACACAGA----- 4763
QY 141 rgGlyArgGlnGlyArgHisHisValGlnGlu-----TyrProValGluPheProA 158
Db 4764 GAAACCCGAAACGGAACCAACAGACAGAAACCAACCAACCAACCAACCAACCAACGCGC 4823

```

[illegible]

```

1 CITY: Washington, D.C.
2 COUNTRY: USA
3 ZIP: 20007-5109
4
5 COMPUTER READABLE FORM:
6 MEDIUM TYPE: Floppy disk
7 COMPUTER: IBM PC compatible
8 OPERATING SYSTEM: PC-DOS/MS-DOS
9 SOFTWARE: Patent in Release #1.0, Version #1.25
10
11 CURRENT APPLICATION DATA:
12 APPLICATION NUMBER: US/08/173,508
13 FILING DATE: 23-DEC-1993
14 CLASSIFICATION: 435
15 ATTORNEY/AGENT INFORMATION:
16 NAME: BENT, Stephen A.
17 REGISTRATION NUMBER: 29,768
18 REFERENCE/DOCKET NUMBER: 18740/125/CACO
19 TELECOMMUNICATION INFORMATION:
20 TELEPHONE: 202 672 5300
21 TELEFAX: 202 672 5399
22
23 TELEX: 904136
24
25 INFORMATION FOR SEQ ID NO: 7:
26 SEQUENCE CHARACTERISTICS:
27 LENGTH: 1820 base pairs
28 TYPE: nucleic acid
29 STRANDEDNESS: double
30 TOPOLOGY: linear
31
32 FEATURE:
33 NAME/KEY: CDS
34 LOCATION: 104..1720
35
36 FEATURE:
37 NAME/KEY: sig_peptide
38 LOCATION: 104..244
39
40 FEATURE:
41 NAME/KEY: mat_peptide
42 LOCATION: 245..1720
43
44 US-08-173-508-7
45
46
47 Alignment Scores:
48 Pred. No.: 0.0134 Length: 1820
49 Score: 127.50 Matches: 145
50 Percent Similarity: 33.95% Conservative: 75
51 Best Local Similarity: 22.38% Mismatches: 251
52 Query Match: 2.78% Indels: 182
53 Gaps: 30
54
55 US-09-720-086-6 (1-859) x US-08-173-508-7 (1-1820)
56
57 QY 37 AAlAProSerPProProValLeuGlAlAlIECyEThrGluProValCyThrPro-GluTh 56
58 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
59 135 GCACCGGGCGGACCCGTTTCGGGGCCACGCTGCACCGCGCGGCTGCTGCCACCGCCT 194
60
61 QY 56 FArgGIVArGArGserSerSerArgLeuSerLyArGGLuValSerSerLeuLeuAenTy 76
62 195 GCTGGGCGCGG-----GGGCGGCTGCAGCTCGCGCGGATCCCCCGCGCCAAAGCGG 245
63
64 QY 76 FThrGlnAspMetThrGlyAspGlyAspArgAspAspGluValAlAspAspGlyAsnGlySe 96
65 ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
66 246 CCGGGCGACGAGAGCGGCGCCACGGCGACCTGACCCCCCTGCCGAA-----GGCCAGCG 299
67
68 QY 96 FAspIleuMetProLySleuThrArGGLuThrLyAspThrArGThrArgSerGluSe 116
69 ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
70 300 CCGCGGACGCTCCCGTACTACGACGAGAAAGCTCGGCTGGCGCGACTGGG-----CG 353
71
72 QY 116 FProAlaValArGThrArGHisSerAsnGlyThrSerSerLeuGluArGInArg----- 134
73 334 TCCCGGGCTTCAGTGGCGCACCATGAAAGGCCCGGCTGACTACGCCAAAGCCCGCGCAGC 413
74
75 QY 135 -----AlaSerProArgI 139
76
77 Db 414 GCGAGCTCCGGCTCGCGGTGGCCCGCAAGAAAGCCACAGGGGCGGGCGAGACGCTCGGCT 473
78 :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
79 139 eThArGIGlyArGInGlyArGHisHisValGInGluTyTrProValGluPheProAlaTh 159
80 :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

Db 474 CGCTGCTGCTCAACCCGGCGGAC - CGGGCGGCTCGCGATCGGCTACCTCCAGCATC 532  
Qy 159 rArgSerArgArgArgAlaSerSerSerAlaSerThrProTrp-----SerSerPr 177  
Db 533 CGGGGATCGGCTACCCCGGGAAGTCCGGCCAGTACGACATGTTGGCGGTGACCC 592  
Qy 177 oAlaSerValaAepPheMetGluGluValThrProLysSerValSerThrProSerVala 197  
Db 593 CGGG-----CGGTGGCCCGCACTGAA 613  
Qy 197 pleuSerGlnAepGlyAepGlnGluGlyMetAspThr-----ThrGlnValaAepAlaGlu 216  
Db 614 CCGGTGAGTGCCTG - GACGGCCCGAGATGGACCGGTACACCGCCACCGACGTCACCC 672  
Qy 216 rArgAepGlyAepSerThrGluTyrGlnAepAep-----LysGluPheGlyIleGly-- 233  
Db 673 GGACGACGGCGGAGCGGAGCTGGTTCGACGCTACAGGAGTTCGCGGAGGCTG 732  
Qy 234 -----AepLeuValTrpGlyLysIleLysGlyPheSerTrpTrpProAlaMetValVa 251  
Db 733 CGGGCGGACGCG-----CGGAGCTGCTGCG 759  
Qy 251 lSerTrpLysAlaThrSerLysArgGlnAlaMetProGlyMetArgTrpValGlnTrpPh 271  
Db 760 CCACGCTCCAGCTCGAGGCGGACGACATGGAGCTCTGCGCGGTG-----CT 813  
Qy 271 eGlyAepGlyLysPheSerGluIleSerAlaAepLysLeuValAlaLysGlyLeuPheSe 291  
Db 814 GGGGACGAGAACTGACCTACGTGGGAGGCTGCTACGCGACCTTCTTGSGC----- 865  
Qy 291 rGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLys----- 307  
Db 866 -----GGACCTACCGCGGTCTGTTCGCCGACCGAGCGGCGCGCTGGT 909  
Qy 308 -----AlaMetTyrHisThrLeuGluLysAlaArgValArgAlaGlyLysThr-- 323  
Db 910 CTTGACGCGCGCATGACCCCTCGTGCCTCCCGCCCGCTGAACTTGGAGCAGACGGA 969  
Qy 324 ----PheSerSerProGlyGluSerLeuGluAepGlnLeuLys---ProMetLeuG 341  
Db 970 GGGCTTCGAGAGCGGCTTCCAGTCTTCGCGAAGGACTCGGTGAAGCAGCGGA -CTGCC 1028  
Qy 341 utrpAlaHisGlyGlyPheLysProThrGlyIleGluGlyLeuLysProAsnLysG 361  
Db 1029 CCTCGCGCACAAAGCACACCC-----CGACCGCTCGCAAGA 1070  
Qy 361 nProValValaAsnLysSerLysValArgArgSerAepSerArgAsnLeuGluProArg 381  
Db 1071 ACTCAAGTC-----CTTCTTCAGCA-----CCTGACG 1100  
Qy 381 gArgGluAsnLysSerArgArgArgThrThrAsnAepSerAlaAlaSerGluSerPro 401  
Db 1101 CGAAGCCCTCGCCCGCGGACCGG -ACGGCCGCAAGCTCACCGAATCCTCGCCACC 1159  
Qy 401 oProLysArgLeuLysThrAsnSerTyrGlyLysAepArgGlyGluAepGluGluSe 421  
Db 1160 ACCGGCGTG -ATCCGCGCGATG-----TACGACGAGGCGCGCTGGCAGCAGCT 1206  
Qy 421 rArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAepArgCysLe 441  
Db 1207 GCGCGAGTCCCTCACCTCGCGCATCAAGGAGGAGGACGCT----- 1246  
Qy 441 uSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyGlyLeuCysG 461  
Db 1247 -----GCGGCGCTGCTGATCCT 1263  
Qy 461 nSerCysArgAepArgPheLeuGluLeuPheTyrMetTyrAepGluAepGlyTyrGlnSe 481  
Db 1264 CTCC-----GACAGCTACTACGAGCGCGGCGGCGCGCC----- 1297  
Qy 481 rTyrCysThrValCysGlyGluArgGluLeuLeuLysSerAsnThrSerCysCy 501  
Db 1298 -----GGCTACAGCAACTGATGTTCCGCAACCGCCG----- 1330

Qy 501 eArgCysPheCysValGluCysLeuGluValLeuValGlyAlaGlyThrAlaGluAspAl 521  
Db 1331 -----GTGAACCTGCTGACCTC-----CCCGC 1353  
Qy 521 alysLeuGlnGluProTrpSerCysTyrMetCysLeuPro-----GluArgCysHisGl 539  
Db 1354 CGCCTTCTCTCCCGGACGAGTGGCGGCGCTCCCGACTTCGAGAGCGGTCCCC 1413  
Qy 539 yValLeuArgArgArgLysAspTrpAsnMetArgLeuGlnAspPhePheThrAspPr 559  
Db 1414 GGTCTTGGGAGGGCTCGCTGTCTCTCCCTGACTGCGCGTACTGGCGGTGAAGCC 1473  
Qy 559 aAspLeuGluGluPheGluProProLysLeuTyrProAlaIleProAlaAlaLysArg 579  
Db 1474 C-----ACGGGGAGCGCGCACCGC-----ATCAGGCGCGCGCGCCAC 1512  
Qy 579 gProIleArgValLeu-----Se 585  
Db 1513 CCGCATGCTGTGTGTCGACACCGCGGACCGCGCCCTACCGCTGGGCGGAGGC 1572  
Qy 585 rLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeuGlyIleLysValGl 605  
Db 1573 CCTCTCCGACGAGCTCACCTCGCGCACCTCTCTACCTACGAG---GGAGCGGCGCAC 1629  
Qy 605 ulysTyrIleAlaSerGluValCysAlaGluSer-----IleAlaValGl 620  
Db 1630 CGCGTACGGCGCGGAGCTCTCTGATCGATCTCGCATCAACAGTACTGTCGACCG 1689  
Qy 620 yThrValLysHisGluGlyGln 627  
Db 1690 CACCGCGCGGAGGACGCGCAAG 1711

## RESULT 9

US-08-265-310-7  
; Sequence 7, Application US/08265310  
; Patent No. 5856166  
; GENERAL INFORMATION:  
; APPLICANT: Bartfeld, Daniel  
; APPLICANT: Butler, Michael J.  
; APPLICANT: Hadary, Dany  
; APPLICANT: Jenish, David  
; APPLICANT: Krieger, Timothy  
; APPLICANT: Malek, Lawrence T.  
; APPLICANT: Soostmeyer, Gisela  
; APPLICANT: Walczyk, Eva  
; APPLICANT: Krygman, Phyllis  
; APPLICANT: Gaten, Sheila  
; TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED  
; TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND  
; TITLE OF INVENTION: POLYPEPTIDES  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/265,310  
; FILING DATE: 24-JUN-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,508  
; FILING DATE: 23-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Benti, Stephen A.  
; REGISTRATION NUMBER: 29,768



```

REFERENCE/DOCKET NUMBER: 18740/133/CACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1821 base pairs
type: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 104..1720
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 104..244
ATTURE:
NAME/KEY: mat_peptide
LOCATION: 245..1720
US-08-265-310-7

Alignment Scores:
Pred. No: 0.0134 Length: 1821
Score: 127.50 Matches: 145
Percent Similarity: 33.95% Conservative: 75
Best Local Similarity: 22.38% Mismatches: 251
Query Match: 2.78% Indels: 182
DB: Gaps: 30

US-09-720-086-6 (1-859) x US-08-265-310-7 (1-1821)
QY 37 AlaProSerProProValLeuGluAlaLeuCysThrGluProValCysThrPro-GluTh 56
Db 135 GCACCGGGGACCCGTTCCGGGGCCAGCTGCTACCGCGCGCTTCCGACCCGCTT
QY 56 rArgGlyArgArgSerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTy 76
Db 195 GCTGCGCGCGG-----GGGCGGCTGACGCTCCGCGATCCCGCGGGCGCAAGGCGG 245
QY 76 rThGlnAspMetThrGlyAspGlyAspArgAspArgValAspArgGlyAsnGlySe 96
Db 246 CCGGCGGACGAGGCGGCGGCGGACCTGACCCCTCGCCGCA-----GGCCAGCC 299
QY 96 rAspIleuMetProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSe 116
Db 300 CCGCGAGACTGCTCCCTACTACGACGAGAGCTCGCTGGCGGCACTGCGG-----CG 353
QY 116 rProAlaValArgThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArg----- 134
Db 354 TCCGGGCTTCCAGTGGCGGCAACATGAAGGCCCGCTGACTAGCCCAAGCCCGCGACG 413
QY 135 -----AlaSerProArgI 139
Db 414 GCGAGCTCCGCTCGGCTGGCGGCGCAAGGCCACGCGGCGGCGGCAAGCGCTCGGCT 473
QY 139 eThArgGlyArgGlnGlnArgHisHisValGlnGluThrProValGluPheProAlaTh 159
Db 474 CGCTGCTGTCAACCCGGGCGGAC-CGGGCGGCTCGCGGATCGGCTTCCAGCAAGTAC 532
QY 159 rArgSerArgArgArgAlaSerSerSerAlaSerThrProTrp-----SerSerPr 177
Db 533 GCGGGCATCGGCTACCGCGGCAAGGTCCGCGCCGACTACGACATGTGTGGCGGTCAACC 592
QY 177 rAlaSerValAspPheMetGluGluValThrProLysSerValSerThrProSerValAs 197
Db 593 CGGG-----GCGTGCGCGCGAGTAA 613
QY 197 rLeuSerGlnAspGlyAspGlnGluGlyMetAspThr-----ThrGlnValAspAlaGluSe 216
Db 614 CCGCTGAGTGCCTTG-AGCGGGCGGAGATGAGCGCTACGACCGACCGGATCAACCC 672
QY 216 rArgAspGlyAspSerThrGluThrGlnAspAsp-----LysGluPheGlyIleGly-- 233

```

```

Db 673 GAGACGACGGGCGGACGAGCGAGCTGTGTCAGCTTACAGAGAGTTCGCGGAGGCTG 732
QY 234 -----AspLeuValThrGlyLysIleLysGlyPheSerThrTrpProAlaMetVal 251
Db 733 CGGGGCGGACCG-----CGAAGCTGTGCGG 759
QY 251 lSerTrpLysAlaThrSerLysArgGlnAlaMetProGlyMetArgTrpValGlnTrpPh 271
Db 760 CCAGCTTCCACGCTGACAGGGGACCGGACATGAGCTGCTGCGCGCGTG-----CT 813
QY 271 eGlyAspGlyLysPheSerGluLysSerAlaAspLysLeuValAlaLeuGlyLeuPheSe 291
Db 814 GGGCGACGAGAGCTGACTACGAGGAGCTGTACGCGACCTTCTCGGCG----- 865
QY 291 rGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerThrTrpLys----- 307
Db 866 -----CGACCTACGCGCGGTGTCTTCCCGACGAGCGGCGCGCTGCT 909
QY 308 -----AlaMetTrpHisThrLeuGluLysAlaArgValArgAlaGlyThr-- 323
Db 910 CCTGAGCGGCGCGATGACCCCTGCTGCGCGCGCGCTGAACTGAGGACGACGAG 969
QY 324 ----PheSerSerSerProGlyGluSerLeuGluAspGlnLeuLys--ProMetLeuG 341
Db 970 GGGCTTCGACGCGCGTTCCTCCAGCTCTTCGCGAGAGACTGCTGAAGCAGCGCGA-CTGCC 1028
QY 341 uTrpAlaHisGlyLysPheLysProThrGlyIleGluGlyLeuLysProAlaAsnLysG 361
Db 1029 CCGTGGCGGACAGACCAACCC-----CGACCTAGGTTCGCGAAGA 1070
QY 361 nProValAlaAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProAlaArg 381
Db 1071 ACTTCAAGTC-----CTTCTTCAGCA-----CTTGACG 1100
QY 381 rArgGluAsnLysSerArgArgArgThrThrAsnAspSerAlaAlaSerGluSerProPr 401
Db 1101 CGAAGCCCTTCCCGCGGCGGACCGG-ACGGCGGCAAGCTCACGAACTCCTCGCCAC 1159
QY 401 oProLysArgLeuLysThrAsnSerTyrglyGlyLysAspArgGlyGluAspArgGluSe 421
Db 1160 ACCGCGCTG-ATCGCCCGATG-----TACGACAGGCGCGCTGCGACAGCT 1206
QY 421 rArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCysLe 441
Db 1207 GCGGAGTCTCTACCTCGGCGATCAAGAGAGAGAGCT----- 1246
QY 441 uSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyLeuCysG 461
Db 1247 -----GCGGCGCTGCTGATCCT 1263
QY 461 nSerCysArgAspArgPheLeuGluLeuPheTyMetTyAspGluAspGlyTyrglnSe 481
Db 1264 CTCC-----GACAGCTACTACGAGCGGCGGACGAGC----- 1297
QY 481 rTyrcyethrValCysCysGluGlyArgGluLeuLeuLeuCysSerAsnThrSerCys 501
Db 1298 -----GCTACAGCAACCTGATGTTGCCCAAGCGCGCC----- 1330
QY 501 sArgCysPheCysValGluCysLeuGluValAlaGlyAlaGlyThrAlaGluAspAl 521
Db 1331 -----GTGAAGTCTCTGCACTC-----CCGCG 1353
QY 521 aLysLeuGlnGluProTrpSerCysTyMetCysLeuPro-----GlnArgCysHisG 539
Db 1354 GCGCTTCTCTCCCGGAGCGAGTCCGCGAGCCCTCCCGCACTTCAGAGAGCGGTCC 1413
QY 539 yValLeuArgArgArgLysAspTrpAsnMetArgLeuGlnAspPhePheThrThrAspPr 559
Db 1414 GGTCTTGGCGAGGAGCGCTCGCTGCTCTCTTAAGTGGCGGCTAGCGCGGAGAAC 1473
QY 559 oAspLeuGluGluPheGluProProLysLeuTyTrpAlaAlaProAlaAlaLysArgAr 579

```

Db 1474 C-----ACGGGGAGCGCACCGC-----ATCGAGGGCGCGCGCCAC 1512  
Qy 579 gProfileArgValLeu-----  
Db 1513 CCGATCGTGTGGTGGCACACACCGCGACCGCGCCACCCCTACCGCTGGCGCGAGGC 1572  
Qy 585 rLeuPheAspGlylleAlaThrGlyrLeuValLeuValLeuGlylleTysValG1 605  
Db 1573 CTTCTCCGACCAAGCTCACCTCGCGCCACCTCTCTCACTACGAG---GGAGACGGCCACAC 1629  
Qy 605 ulvstYrilleAlaserGluValCysAlaGluSer-----lleAlaValG1 620  
Db 1630 CGGTAGCGCGCGCGCACCTCTCTCATCGACTCGCGATCAACACGTACTCTGTGACCGG 1689  
Qy 620 yThrVallyshHisGluGlyGln 627  
Db 1690 CACCGCCCGGAGGACGGCAAG 1711  
RESULT 10  
US-08-951-742-7  
; Sequence 7, Application US/08951742  
; Patent No. 6127144  
; GENERAL INFORMATION:  
; APPLICANT: Bartfeld, Daniel  
; APPLICANT: Michael J. Butler  
; APPLICANT: Dany Hadary  
; APPLICANT: David Jenish  
; APPLICANT: Tim Krieger  
; APPLICANT: Lawrence T. Malek  
; APPLICANT: Gisela Soostmeyer  
; APPLICANT: Eva Walczyk  
; APPLICANT: Phyllis Krygsmann  
; APPLICANT: Sheila Garven  
; TITLE OF INVENTION: METHOD FOR EXPRESSION OF PROTEINS IN  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; City: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/951,742  
; FILING DATE: 16-OCT-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bent, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 0189740/0140  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1821 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 104..1720  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 104..244  
; FEATURE:

; NAME/KEY: mat\_peptide  
; LOCATION: 245..1720  
US-08-951-742-7  
Alignment Scores:  
Pred. No.: 0.0134 Length: 1821  
Score: 127.50 Matches: 145  
Percent Similarity: 33.95% Conservative: 75  
Best Local Similarity: 22.38% Mismatches: 251  
Query Match: 2.78% Indels: 182  
DB: 3 Gaps: 30  
US-09-720-086-6 (1-859) x US-08-951-742-7 (1-1821)  
Qy 37 AlaProSerProProValLeuGluAlaIleCysThrGluProValCysThrProGluTh 56  
Db 135 GCACCGCGGACCCGTTTCGGCGGCACGCTGCTACCGCGGCTGCTCGCACCGCT 194  
Qy 56 rArgGlyArgSerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTy 76  
Db 195 GCTCGGCGG-----GGGCGGCTGACGCTCCGCGGATCCCGCGGCGCAAGGCGG 245  
Qy 76 rThrGlnAspMetThrGlyAspGlyAspArgAspGluValAspAspGlyAsnGlyse 96  
Db 246 CCGCGCGGACGGGCGGCACGGGACCCCTGACCCCTCGCCGAA-----GGCCACGG 299  
Qy 96 rAspIleLeuMetProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSe 116  
Db 300 CCGCGGAGCTGCTCCCGCTACTACGAGACAGAACTCGGCTGGCGGACTGCGG-----CG 353  
Qy 116 rProAlaValArgThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArg----- 134  
Db 354 TCCCGGGCTTCAGTGGCGGCACCATGAAGCGCCCGCTCGACTACGCCAAGCCCGCGAGC 413  
Qy 135 -----AlaserProArgI1 139  
Db 414 CCGAGCTCGGCTCGGCTGGCGCCCAAGAGCCACGGGCGCGGCAAGCCCTCGGCT 473  
Qy 139 eThrArgGlyArgGlnGlyArgHisValGlnGluTyProValGluPheProAlaTh 159  
Db 474 CCGTCTGGTCAACCCCGGCGGAC-CGGGCGGCTCGGCGATCGGCTACCTCCAGCAGTAC 532  
Qy 159 rArgSerArgArgArgAlaserSerSerAlaserThrProTrp-----SerSerPr 177  
Db 533 CGCGGCGATCGGCTACCCCGCGAAGGTCCGCGCCAGTACGACATGTTGGTGGTCCAGCCCC 592  
Qy 177 oAlaserValAspPheMetGluGluValThrProLysSerValSerThrProSerValas 197  
Db 593 CGGG-----GCGTGGCGCGCGAGTAA 613  
Qy 197 pleuSerGlnAspGlyAspGlnGluGlyMetAspThr---ThrGlnValAspAlaGluSe 216  
Db 614 CCGCTCGAGTGCCTG-GACGGGCGCGAGATGGACGCGTACACCGCGCACCGCTCACCCCC 672  
Qy 216 rArgAspGlyAspSerThrGluTyGlnAspAsp-----LysGluPheGlyIleGly-- 233  
Db 673 GGACGACGGCGGCGAGCGGACGAGCTGGTGCAGCCCTACAGAGAGTTTCGCGAGGGCTG 732  
Qy 234 -----AspLeuValTrpGlyLysIleLysGlyPheSerTrpTrpProAlaMetValva 251  
Db 733 CGGGCGGACCGG-----CCGAGCTGCTGG 759  
Qy 251 lserTrpLysAlaThrSerLysArgGlnAlaMetProGlyMetArgTrpValGlnTrpPh 271  
Db 760 CCACGTCTCCACGGTCCAGCGCGGCGGACGACATGGAGCTCTCGCGCGGTTG-----CT 813  
Qy 271 eGlyAspGlyLysPheSerGluIleSerAlaAspLysLeuValAlaLeuGlyLeuPheSe 291  
Db 814 GGGCGACGAGAAGCTACCTACGTGGGAGCGCTCGTACGGCACCTTCTCGGC----- 865  
Qy 291 rGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLys----- 307  
Db 866 -----GCGACCTACCGCGGCTCTGTTCCCGGACCGGACCGCGGCGGCTGCT 909

QY	308	-----AlMeTyrHisThiHeuGluValAlaArgValaGlyLysThr--	323
Db	910	CTTGACGAGCGGATGGACCCCTGGCTGCGCGCGCGCGCGCTGGACCTGGACGACGAG	969
QY	324	-----PheSerSerProGlyGluSerLeuGluAerGlnLeuLys---ProMetLeuG1	341
Db	970	GGGCTTGACAGCGCGCTTCAGCTCTTCGGAAAGACTCGCGAAAGACCGGACCTGCC	1028
QY	341	UTPrAlaHisGlyGlyPheLysProThrThrglyIleGlyIleuLysProAsnLysGly	361
Db	1029	CCCTCGGCGACAAAGACACACACCC-----CGACCGAGTCGGCAAGA	1070
QY	361	nProValValAsnLysSerLysValAlaArgAserAspSerAlaAsnLeuGluProAlaArg	381
Db	1071	ACCTCAAGTC-----CTTGTTCGACGA-----CTCGAGCG	1100
QY	81	gArgGluAsnLysSerArgArgThrgThrHisAsnAspSerAlaAlaSerGluSerProPr	401
Db	1101	CGAAGCCCTGCGCGCGCGGAGGAGCGG-AGGCGCGCAACTGACCGAATCCTCGCACCC	1159
QY	401	oProLysArgLeuLysThrAsnSerThrglyGlyLysAspArgGlyGluAspGlyLysE	421
Db	1160	ACGGCGCGG-ATCGCGCGGATG-----TACGACAGAGGCGCGCTGGACGACACT	1206
QY	421	rArgGluArgMetAlaSerGluValAlaThrAsnLysGlyLeuAsnLeuGluAspArgCysLe	441
Db	1207	GCGGAGATGCTCTCACTCGCGCGATCAAGAGGAGAGGAGCGGT-----	1246
QY	441	uSerCysGlyLysLysAsnProValSerPheHisProLysPheGlyGlyGlyLeuCysGly	461
Db	1247	-----GCGGCGCTGCGTATCT	1263
QY	461	nSerCysArgAspArgPheLeuGluLeuPheThrglyMetThrgArgLysAspGlyThrgLysE	481
Db	1264	CTCC-----GACAGCTACTACAGCGCGGAGCGCGACGCG-----	1297
QY	481	rTyrCysThrValCysCysGlyLysArgLysLeuLeuLeuCysSerAsnThrSerCysCys	501
Db	1298	-----GGCTACAGCAACTGATGTTGCGCAACGCGCGC-----	1330
QY	501	sArgCysPheCysValGlyCysLeuGluValLeuValGlyAlaGlyThrAlaGluAspAl	521
Db	1331	-----GTGAACGCTCTCGACCTC-----CCGCG	1353
QY	21	AluLysLeuGlnGluProThrPheSerCysThrglyMetCysLeuPro-----GlnArgCysHisG1	539
Db	1354	GCGCTCTCTCTCTCTCGGACAGAGTGCGGACAGCGCTCTCTCGGACTTCGAAAGCGCTCC	1413
QY	539	yValLeuArgThrgThrgLysAspTrpAsnMetArgLeuGlnAerPheThrThrAspPr	559
Db	1414	GGTCTTGAGGAGGCGCTCGCTGATCTCTCGGAACTCGGGTACTGGCGGTAAAC	1473
QY	559	oAspLeuGlnGluPheGluProProLysLeuThrgProAlaIleProAlaIleuLysArg	579
Db	1474	C-----ACGGGAGGAGCGGACGCG-----ATCGAGGCGCGCGGCGCAC	1512
QY	579	gProIleArgValLeu-----Se	585
Db	1513	CCGATGCTGCTGGTCTGGCGACACCCGCGACCCGCGACCCCTACCGCTGGGCGGAGCG	1572
QY	585	rLeuPheAspGlyLeuAlaThrGlyThrgLysValLeuLysGlyLeuGlyIleLysValG1	605
Db	1573	CTCTCCGACGACGCTACCTCTCGGACACTCTCTCACACTACAG--GGAAGCGGCAAC	1629
QY	605	uLysThrIleAlaSerGluValCysAlaGluSer-----IleAlaValG1	620
Db	1630	CGCGTACGCGCGGCGAGCTCTGATCACTCGGATCAACAGCTACTCTGACGCG	1689
QY	620	yThrValLysHisGlnGlyGln	627
Db	1690	CACGCGCCCGGAGGAGCGCAAG	1711

```

RESULT 11
US-08-973-675-1
Sequence 1, Application US/08973675
Patent No. 5985283
GENERAL INFORMATION:
APPLICANT: HATEBOER, GUS
APPLICANT: BERNARDS, REME
TITLE OF INVENTION: ADENOVIRUS E1A-ASSOCIATED PROTEIN BS69,
INHIBITOR OF E1A-TRANSACTIVATION
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHAYE P. C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,675
FILING DATE: 12-DEC-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: WILSON, MARY J.
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 620-31
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2623 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 150..1835
US-08-973-675-1

Alignment Scores:
Pred. No.: 0.0278 length: 2623
Score: 127.00 Matches: 67
Percent Similarity: 30.42% Conservative: 41
Best Local Similarity: 18.87% Mismatches: 113
Query Match: 2.77% Indels: 134
DB: 2 Gaps: 12

US-09-720-086-6 (1-859) x US-08-973-675-1 (1-2623)
QY 234 AspleuValItrpGlyLySleLySgLyPheSerTrpTrpProAlaMetValValSerTrp 253
:::|||||:::|||||:::|||||:::|||||:::
Db 873 GAGCTGGTTGGCGTAAATGAAGGTTTGGGTTTGGCCAGCCAAAGTCATGAG--- 929
QY 254 LysAlaThrSerLySgAlaGlnAlaMetProGlyMetArgTrpValGlnTrpPheGlyAsp 273
Db 929.----- 929
QY 274 GlyLySPheserGluIleSerAlaAspLySLeuValAlaLeuGlyLeuPheSerGlnHis 293
930-----AAAGAAAGCATCAAGTCGACGTTGCTTTGGCCACAC 971
Db 294 PheAsnLeuAla-----ThrPheAsn----- 300
972 CACAGAGGGGCTGATTCTTCTGAAACATTCGAAGATATACACAGTCAACATTCATCGG 103
QY 301-----LysLeuValSerTyArgLySAlaMetTyHis---ThrLeuGlnLys 315
:::|||||:::|||||:::

```

Db 1032 CTGACGCTGAAGCGCAGTATGGTTGGAAAAAGGCGCTGTGATGAGCTGGAGCTGCATCAG 1091  
Qy 316 AlaATgValaTgAlaGlyLys----- 322  
Db 1092 CGTTTCTACGAGAAGGAGATTTTGGAAATCTAAGAATGAGGACCGAGGTGAGGAAGAG 1151  
Qy 323 -----ThrPheSerSerProGlyGluSerLeuGluAspGlnLeuLysProMet 339  
Db 1152 GCAGATCCAGTATCTCTCCACAGTAATGAGCAGCTAAAGTCACTCAAGAACCAAGA 1211  
Qy 340 LeuGluTTPAlaHisGlyLysProThrGlyLysGluGluGluGluGluGluGluGluGlu 359  
Db 1212 GCAAAG-----AAAGGACGAGCTAATCAAGTGTGGAGCCCAAAAAG 1253  
Qy 360 LysGlnProValValAsnLysSerLysValArgSerAspSerArgAsnLeuGluPro 379  
Db 1254 GAAGAACCCAGACGCTGAACAGAGCAGTAAGTTCTAGCCAGGAATAACCCAGATGCCT 1313  
Qy 380 ArgArgArgGluAsnLysSer----- 386  
Db 1314 CAGCCCATCGAAAAAGTCTCGGTGTCACTCAGACAAAGAAAGTTAAAGTGCCTTTCACCA 1373  
Qy 387 -----ArgArgArgThrAsnAspSerAlaAlaSerGluSerProPro 401  
Db 1374 AGAATGCTGCATCGGAGCACCAGACCCACCAACGCGGGTGTGTGAGAGCATGTGCCAT 1433  
Qy 402 ProLysArgLeuLysThr-----AsnSerTyrGlyLysAspArgGlyGlu----- 417  
Db 1434 GACAAATACCAAGATCTTCAATGACTTC-----AAAGACCGGATGAAGTCGGACCCAC 1487  
Qy 418 -----Asp 418  
Db 1488 AACCGGGACAGACGCTGTGTCCGAGAAGCTCTGGAGAAGCTCGCTTCTGAATGGAA 1547  
Qy 419 GluGluSerArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAsp 438  
Db 1548 GAAGAAAGAGCAAGCTGTAAATAAGCTGTAGCCACATCGAGGGTGAGATGGACAGA 1607  
Qy 439 ArgCysLeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyGly 458  
Db 1608 AAATGTAGCAAGTAAGGAAAAAG----- 1631  
Qy 459 LeuCysGlnSerCysArgAspArgPheLeuGluLeuPhe----- 471  
Db 1632 -----TGTAGGAGGAATTTGTAGAGAAATCAAGAAAGCTGGCAACACAGCAC 1679  
Qy 472 --TyrMetTyrAspGluAspGlyTyrGlnSerTyrCysThrValCysCysGluGlyArg 490  
Db 1680 AAGCAACTGATTTCTCAGACCAAGAGAGCAGTGTGTCTACAACTGTGAGGAGGAGGCC 1739  
Qy 491 GluLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCys 505  
Db 1740 ATGTACCAGTCTGCTGCAACACATCTTACTGTCTCATCAAGTGC 1784

## RESULT 12

US-09-453-702B-96/c

Sequence 96, Application US/09453702B

Patent No. 6365723

GENERAL INFORMATION:

APPLICANT: Blattner, Frederick R.

Burland, Valerie

Ferna, Nicole T.

Plunkett, Guy

Welch, Rod

TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESS: Quarles &amp; Brady

STREET: 1 South Pinckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53701-2113

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/453,702B

FILING DATE: 03-Dec-1999

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/110,955

FILING DATE: 04-Dec-1998

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27386

REFERENCE/DOCKET NUMBER: 960296.95017

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 251-5000

TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 96:

SEQUENCE CHARACTERISTICS:

LENGTH: 34063

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 96:

US-09-453-702B-96

Alignment Scores: 2.99 Length: 34063

Score: 125.50 Matches: 58

Percent Similarity: 40.69% Conservative: 36

Best Local Similarity: 25.11% Mismatches: 88

Query Match: 2.73% Indels: 49

DB: 4 Gaps: 11

US-09-720-086-6 (1-859) x US-09-453-702B-96 (1-34063)

Qy 573 lleProAlaLysArg-----ProIleArgValLeuSerLeuPheAspGlyIle 590  
Db 31583 ATAGAACTAATAAGAGACGTTGCCAACTATATGAATGATATAGATTTGTTTCTGCGCTT 31524  
Qy 591 AlaThrGlyTyrLeuValLeuLysGluLeuGlyLysValGluLysTyrIleAlaSer 610  
Db 31523 GGAGTCTAAGTCTTGTGCTGCAGCTGCTGATTTGATGTTAA-----ATGCGAGTT 31470  
Qy 611 GluValCysAlaGluSerIleAlaValGlyThrValLysHisGluGlyGlnIleLysTyr 630  
Db 31469 GAATTTGATCAACATGCTATTAACTACCCCAATTAATTTCCAGAAGTTTGCACGTC 31410  
Qy 631 ValAsnAspVal-----ArgLysIleThrLysLysAsnIleGluLysTyrGly 646  
Db 31409 CAGAAGATGTTTCTTACTAATGCAGAAATATTAGGGTTTTTTTAAACCATATG 31350  
Qy 647 ProPheAspLeuValIleGlySerProCysAsnAspLeuSerAsnVal----- 663  
Db 31349 CCCATAGATGTTATTATGGCGTCTCCGTCGAAGATTTAGTTCAATAGCAAGGGG 31290  
Qy 664 AsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPheGluPheTyrHis 683  
Db 31289 AATCCTGAT-----GATAGCAGGAATCAGCTTTACATGCAATTTCTACCGT 31245  
Qy 684 LeuLeuAsnTyrThrArgProLysGluGlyAspAsnArgProPheThrPheMetPheGlu 703  
Db 31244 TTAGTATCA-----GAATTACAGCAATTTCTTTTGGCAGAA 31206  
Qy 704 AsnValValAlaMetLysValAsnAspLysLysAspLysSerArgPheLeuAlaCysAsn 723  
Db 31205 AATGTTCCAGGTATT--ATGCAAGAGAAATATTCTGCATAGAAATAAAGCAATTTAAT 31149  
Qy 724 ProVal-----MetIleAspAlaIleLysValSerAlaHis----- 736

Db 31148 TTGGTAGCGGTGATATGATATCTTGATCCCATCAAGTAAAGCATCTGATTATAGT 31089  
Qy 737 -----ATGAAATGTYTPhetTryptGlyAsnLeuProGlyMetAsnArgProVal 752  
Db 31088 GCTCCAACTATTCGAAACTAGATATTTT-----ATC 31056  
Qy 753 MetLaserLysAsnAspLysLeuGluLeuGluAspCysLeuGluPheSerArgThrAla 772  
Db 31055 GGTTAAATAATCATGAACTGATATTTTCAGAT-----GAAGTATTT 31011  
Qy 773 LysLeuLysLysValGlnThrIleThrThrLys 783  
Db 31010 ATGCCTAAATGATGATTCGCGTTACTGTAA 30978  
RESULT 13  
US-09-103-840A-1  
Sequence 1, Application US/09103840A  
Accession No. 6294328  
LOCAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1  
Alignment Scores:  
Pred. No.: 1,11e+04 Length: 4411529  
Score: 125.50 Matches: 156  
Percent Similarity: 30.39% Conservative: 61  
Best Local Similarity: 21.85% Mismatches: 245  
Query Match: 2.73% Indels: 258  
Gaps: 36  
US 20-086-6 (1-859) x US-09-103-840A-1 (1-4411529)  
Qy 60 ArgSerSerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTyThrGlnAsp 79  
Db 1605907 CGACGAGCGCAATAGCTGCTACCGCGAGAGTTACTGCGGTAAACCGCAG 1605966  
Qy 80 MethrGlyArgGlyArgArgAspAspGluVal-----Asp 91  
Db 1605967 CTACGTCGCCACCGCGCGAGATGCTGCACCGCAACACCATTCATACCGGGGAT 1606026  
Qy 92 AspLysAsnGlySerAspLysLeuMetProLysLeuThrArgGluThrLysAspThrArg 111  
Db 1606027 CCAAGCGCATGCACT----- 1606041  
Qy 112 ThrArgSerGluSerProAlaValArgThrArgHisSer----- 124  
Db 1606042 ATCCGACGACAAATCTCGACATCCGATCCGCGCTTCCGGGTGAGATGCGCTGAGGT 1606101  
Qy 125 -----AsnGlyThrSerSerLeuGluArgGlnArgAlaSerProArgGlyLeuThrArg 141  
Db 1606102 CTGCGCGTGGATGGACCGCGCGGTCTCCCGCCAAACAAATAGTGTCTCGTAAACCGCG 1606161  
Qy 142 GlyArgGlnGlyArgGlnHisValGlnGluTyrProValGluPheProAlaThrArgSer 161  
Db 1606162 GTCCGTTCAATCGCTGGGCAC-----AATCGGTGT 1606191  
Qy 162 ArgArgArgAlaSerSerSerSerLysSerThrProTyrSerSerProAlaSerValAsp 181

Db 1606192 CGTGACCTTCGCTGCTGCGCGCAATATGCTCTCCGACGATTCGGCGCT----- 1606239  
Qy 182 PheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGlnAsp 201  
Db 1606240 -----AAGCTTGGCCACAGCTCAACACCGCACGACGAA-----ACCCAGCGC 1606281  
Qy 202 GlyAspGlnGluGlyMetAspThrThrGlnValAspAlaGluSerArgAspGlyAspSer 221  
Db 1606282 GGTCCCGCAGAGGGGGG-CGGACCGCGGTATACCAATTCGGCTGCGTC-GGTTCTAGT 1606339  
Qy 222 ThrGluTyrGlnAspAspLysGluPheGlyIleGlyAspLeuValTryptGlyLysIle--- 240  
Db 1606340 -----TGCGGCTATATCATCCGTTCCACAGGGGTTGGTCGA 1606375  
Qy 241 -----LysGlyPheSer-----TryptProAlaMetValAspSer 252  
Db 1606376 AGGTCCGTATATGCTGCTTTCGGGTGCGCAAGATGCTGGCGGACACCGCTTCGATT 1606435  
Qy 253 TryptLysAlaThrSerLysArgGlnAlaMetProGlyMetArgTryptValGln----- 269  
Db 1606436 TGGCCAGCTCTCGAGCGCGCTGAGCGAGGCCACCGCGCGCGCTATCCACACAC 1606495  
Qy 270 -----TryptPheGlyAspGlyLysPheSerGluIleSerAlaAspLysLeuValAlaLeu 287  
Db 1606496 AAGTACTGGCC-CCGCGCGCC-----GATGAGTGTGCGCGGCATCCGCGAGTTGTT 1606548  
Qy 288 GlyLeuPheSerGlnHisPheAsnLeuAlaThr----- 298  
Db 1606549 GGTCCGACGCGCCAAAGATTTCAAGGCTCAGCGCCACGACATCCGCTTATGACCGG 1606608  
Qy 299 PheAsnLysLeuValSerTyrArgLysAlaMetTyrHisThrLeuGluLysAlaArgVal 318  
Db 1606609 TTCTCGCGGCGCTTAAGCGCGCGAGCGGCTGTATGTGACCGCCAGCGCGCAACGC- 1606667  
Qy 319 ArgAlaGlyLysThrPheSerSerSerProGlyLysLeuGluAspGlnLeuLysPro 338  
Db 1606668 CGCGCTGGT-----GGACACCGCGCGCCACCGCGCGTGGAGTT 1606706  
Qy 339 MetLeuGluTryptAlaHisGlyGlyPheLysProThrGlyIleGlu----- 353  
Db 1606707 GGGGTACGATGGCGCACCGCGCT-----GATTCTGGGCTCCACCGGAACCCCGCAGC 1606760  
Qy 354 -----GlyLeuLysProAsnLysLysGlnProValAlaAsn--- 365  
Db 1606761 GCCCTTGACTACATGACGAGAGTCTACGACCGCTCATCGACCCACACTCTGGGCTA 1606820  
Qy 366 -----LysSerLysValArgArgSerAspSerArg 375  
Db 1606821 TGGCTTTTCCGACCTGTACACCGCCCGCAGTTTCAGCCGTGACCGGCATCCCGACCT 1606880  
Qy 376 AsnLeuGluPro-----ArgArgArgGluAsnLysSerArgArgThrThrAsnAsp 393  
Db 1606881 GACCTACGACCAATCGGTGCGCCGAGCGCGCGCTATCTTACACACGC-----GAT 1606931  
Qy 394 SerAlaLysSerGluSerProPro----- 401  
Db 1606932 CATCGACGAAGTGGCGCGCGGCAATACGTTGAGTGGTTCCTCCGAGGGCGGCTC 1606991  
Qy 402 -----ProLysArgLeuLysThrAsnSerTyrGlyLysAspArgGlyGluAsp 418  
Db 1606992 GGTCCGACCTCGTAATGAGCCATCTGGCAACCTGCGCGCGCGGTGCGCGGAGTCC 1607051  
Qy 419 GluLysSerArgGluArgMetAlaSerGluValThrAsnLysGlyAsnLeuGluAsp 438  
Db 1607052 GATCAAGCTCTGTTGATGTGGGCAACCCCAACACCAACG----- 1607099  
Qy 439 ArgCysLeuSerCysGlyLysLysAsnProValSerPheHis---ProLeuPheGluGly 457  
Db 1607100 -----GGGACCTCTGCGCGGTTCCGGGTCGTACTGACGTGCGCTCGG 1607144  
Qy 458 GlyLeuCysGlnSerCys-----ArgAspArgPheLeuGluLeuPhe 471  
Db 1607145 CTTGACGTTCAACGCTGCGACCCCGGACACCGACTACGCGGACCCAC----- 1607189

QY 472 TyrMetTyrAsp----- 475  
Db 1607190 CATTTACACGACCAATAGCAGCGCTTTCCGACTTCCGAAAGTACCCGCTCAACATCCT 1607249  
QY 476 -----GluAspGlyTyrGlnSerTyrCyThrValCysCysGluGly---- 489  
Db 1607250 GCGGACGTCACGCGCTGCTGGGTATTTACTATTGCG-ACAGCTTGATTACGGGCTCA 1607308  
QY 490 -----ArgGluLeu----- 492  
Db 1607309 CGCCCGACGAGGTGCGCTTCGGGTATCGTCGCGGTGCTTCGCCGAGACCAACACCA 1607368  
QY 493 ---LeuLeuCysserAsnThrSerCysCysArgCysPheCysValGluCysLeuGluVal 511  
Db 1607369 CCTATATTCTGCTTCCCAACAGAGATCTGCGCGTGC---TGCAGCGCTGCGCGGTATTG 1607425  
QY 512 LeuValGlyAlaGlyThrAlaGluAspAlaLysLeuGlnGluProTrpSerCysTyrMet 531  
Db 1607426 TGC-----CCGAGCGCTGCTGGATC 1607446  
QY 532 CysLeuProGlnArgCysHisGlyValLeuArgArgLysAspTrpAsnMetArgLeu 551  
Db 1607447 TCATCGAGCCAGACTGC-----GCGCATCATCAATTGGGTATG----- 1607488  
QY 552 GlnAspPhePheThrAspProAspLeuGluGluPheGluPro---ProLysLeuTyr 570  
Db 1607489 -----ACGACACCGGATACCGCATGTTCCGACCCCGCGCGCA-CTGTC 1607532  
QY 571 Pro-----AlaIlePro-----AlaAlaLysArgArgPro 580  
Db 1607533 CCGGTGCACATGCACCCGATCGAGTCCGCCACAGATAGGCGCTGCGATCGGTGGTCCG 1607592  
QY 581 IleArgValLeuSerLeuPheAspGly----- 589  
Db 1607593 CTCACGCCCTG-----GATGGCTTCTCGACACCGGTATCAACAGATCAACTCAAT 1607643  
QY 590 -----IleAlaThrGly---TyrLeuValLeuLysGluLeuGlyIleLysValGluLys 606  
Db 1607644 CCGTCTGTACACGTCGGGATCTATACAGCCGGGTGCTGAGCTGCGTGGCGCGCGCCG 1607703  
QY 607 TyrIleAlaSerGluValCysAlaGluSerIleAlaValGly 620  
Db 1607704 TACGGTCTCCCGAGCGGTACCAATGCCATTTTATTGGG 1607745  
  
RESULT 14  
US-09-254-325-1  
Sequence 1, Application US/09254325  
Patent No. 6090607  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: ENHANCED EXPRESSION OF  
TITLE OF INVENTION: PROTEOLYTIC ENZYMES IN KOJI MOLDS  
NUMBER OF INVENTIONS: 17  
NUMBER OF SEQUENCES: 17  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30 (BPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/254,325  
FILING DATE:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4657 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1189..1604  
FEATURE:

```

Db 2827 -----GACGGGTTCTCG----- 2838
Qy 258 LysArgGlnAlaMetProGlyWmetArgTyrValGlnTyrPheGlyAspGlyLysPheSer 277
Db 2839 -----TCGGGATGAGTGG-----GATGGGACGTTCCCG 2868
Qy 278 GluIleSerAlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAla 297
Db 2869 -----GCG 2871
Qy 298 ThrPheAsnLysLeuValSerTyrArgLysAlaMetTyrHisThrLeuGluLysAlaArg 317
Db 2872 TCCTTCCATTCGCTGCGCCGCTTTCGCCCTCA-----CATCGCAAGCAT 2916
Qy 318 ValArgAlaGlyLysThrPheSerSerSerProGlyLysLeuGluAspGlnLeuLys 337
Db 17 GTTACCATCGGGTCCACG-----GACATGATGCAC 2946
Qy 338 PrometLeuGluTyrAlaHisGlyLysPhe-----LysProThrGlyTleGluLysLeu 355
Db 2947 ACCCCGAGAGATGGAATCACGGTGGAGTTTGGGTGGACTCATGGG----- 2994
Qy 356 LysProAsnLysLysGlnProValValAsnLysSerLysValArgArgSerAspSerArg 375
Db 2995 -----TCGGTGGCTTCGGTCACTGAGTGCGC-----AACCGA 3027
Qy 376 AsnLeuGluProArgArgArgGlu----- 383
Db 3028 GAGCAGGACCTCGCCGCGCAGAGATTGCGCGCACACGCTCACCCCAATACGCCACAG 3087
Qy 384 -----AsnLysSerArgArgArgThrThrAsnAspSerAlaHisSerGluSerProPro 401
Db 3088 CTGTGGCCCAAGATGACACTATATACATACACTCATACCTCCCTCAATACGCCG 3147
Qy 402 ProLysArgLeuLysThrAsn-----SerTyrGlyGlyLysAsp 414
Db 3148 CCCGAGTCCGCCCTGAGCAGCCGACGTCCTCCGCCCGCGGCGGAGCGAGCAAG 3207
Qy 415 ArgGlyGluAspGluGluSerArgGluArgMetAlaSerGluValThrAsnAsnLysGly 434
Db 3208 AACGGCGAC-----CAAGCGACG 3225
Qy 435 AsnLeuGluAspArgCysLeuSerCys-----GlyLysLysAsn 447
Db 326 AACGAGCCGACCACTGACCAACTGCTTCACTCAACCACTCCGCTGGCGTGGGAAC 3285
Qy 448 ProValSerPheHisProLeuPheGluGly--GlyLeuCysGlnSerCysArgAspArg 466
Db 3286 CCA-----GAGGGCCAGCCGACGTCGCAATGCGCTGC---GGGTTG 3321
Qy 467 PheLeuGluLeu 470
Db 3322 TTTTGAATTTG 3333

RESULT 15
US-09-079-415-5
Sequence 5, Application US/09079415
Patent No. 6013452
GENERAL INFORMATION:
APPLICANT: Christensen, Tove
APPLICANT: Lehmebeck, Jan
TITLE OF INVENTION: A Fungus Wherein The area, pepc and/or
TITLE OF INVENTION: pepc Genes Have Been Inactivated
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESS: No. 60134520 No. 6013452disk of No. 6013452ch America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/079,415
FILING DATE: 14-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 4657,204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 5643 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Aspergillus oryzae
STRAIN: IFO4177
FEATURE:
NAME/KEY: Intron
LOCATION: 2701..2769
NAME/KEY: CDS
LOCATION: join(2282..2700, 2770..4949)
US-09-079-415-5

Alignment Scores:
Pred. No.: 0.16 Length: 5643
Score: 125.00 Matches: 101
Percent Similarity: 30.56% Conservative: 53
Best local Similarity: 20.04% Mismatches: 142
Query Match: 2.72% Indels: 208
Gaps: 24

US-09-720-086-6 (1-859) x US-09-079-415-5 (1-5643)
Qy 9 AsnGluGluGluGluAlaSerGlyTyrGluGluCysIleIleValaAsnGlyAsnPheSer 28
Db 3419 AGCGAGTCTCCGATGGCTCCGGTCATCCG-----TTGCG 3454
Qy 29 AspGlnSerSerAspThrLysAspAlaProSerProProValLeuGluAlaIleCysThr 48
Db 3455 AACCTCATTCGCATACC-----CCGGTGGCTTCGCTCCCTCACTCG 3496
Qy 49 GluProValCysThrProGluThrArgGlyArgArgSerSerSerArgLeuSerLysArg 68
Db 3497 ACCGATTTCTTCTCTCCACCGCCATCAGGCTACGACGATCCAGCGATCC----- 3544
Qy 69 GluValSerSerLeuLeuAsnTyrThrGlnAspMetThrGlyAspGlyAspArgAspAsp 88
Db 3545 -----ACGCGCGAGCCCACTTACGACGCGGAGACCATTCGTT 3580
Qy 89 GluValaAspAspGlyAsnGlySerAspIleLeuMetProLysLeuThrArgGluThrLys 108
Db 3581 TATTTGATATGCGCGTGGG----- 3601
Qy 109 AspThrArgThrArgSerGluSerProAlaValArgThrArgHisSerAsnGlyThrSer 128
Db 3602 GACGCGCCACCCAGCGCGCATTCGCACTATATTTCGATCCGCTCAACTGTTCTGCT 3661
Qy 129 SerLeuGluArgGlnArgAlaSerProArgIleThrArgGlyArgGlnGlyArgHisHis 148
Db 3662 TCGCTGCAG-----CTTCGG----- 3676

```

```
Qy 149 valGlnGluTyProValGluPheProAlaThrArgSerArgArgAlaSerSer 168
Db 3677 -----TATATGTTCCACCAAGAACACCATGACAGCGGCGATTG 3715
Qy 169 SerAla-----SerThrProTrpSerSerProAlaSerValAppPheMetGlu 184
Db 3716 TCACGCGTGCATCCCGGAGTACCCCAATCCCGCCGCAACATGTCGACCCCACTCAG 3775
Qy 185 GluValThrProLysSerValSerThrProSer-----ValAsp 197
Db 3776 GTGTTGAACGCCACCAATATTCTCGACCGCAATCCCACTCCCACTCCCGGCCATGTTTCA 3835
Qy 198 LeuSerGlnAspGlnGluGlyMetAspThrThrGlnValAspAlaGluSerArg 217
Db 3836 TTGAGAGCCGATTACAGTAACGAGATGACGATGTCATCATCTCCGAGCGGCTGGT 3895
Qy 218 AspGlyAspSerThrGluTyGlnAspAspLysGluPheGlyLeGlyAspLeuValTrp 237
Db 3896 CTGCGCATCCGACGAAATATGGGACGAG----- 3925
Qy 238 GlyLysIleLysGlyPheSerTrpTrpProAlaMetValSerTrpLysAlaThrSer 257
Db 3926 -----GACGGGTCTCG----- 3937
Qy 258 LysArgGlnAlaMetProGlyMetArgTrpValGlnTrpPheGlyAspGlyLysPheSer 277
Db 3938 -----TCGGGATCGACGTGG-----GATGGGACGTTCCCG 3967
Qy 278 GluIleSerAlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAla 297
Db 3968 -----GGC 3970
Qy 298 ThrPheAsnLysLeuValSerTyArgLysAlaMetTyHisThrLeuGluLysAlaArg 317
Db 3971 TCCTTCATTCGTCGGGGCTTGGCCCTCAA-----CATCGCAAGCAT 4015
Qy 318 ValArgAlaGlyLysThrPheSerSerSerProGlyGluSerLeuGluAspGlnLeuLys 337
Db 4016 GTTACCATCGGTCCAG-----GACATGATGGAC 4045
Qy 338 ProMetLeuGluTrpAlaHisGlyGlyPhe-----LysProThrGlyLeuGluGlyLeu 355
Db 4046 ACCCCGAGGAGTGGNATCAGTGGCAGTTGGGTGGGACTCATGGG----- 4093
Qy 356 LysProAsnLysLysGlnProValValAsnLysSerLysValArgArgSerAspSerArg 375
Db 4094 -----TCGGTGGCTTCGTCAGTCAGTGGCG-----AACCGA 4126
Qy 376 AsnLeuGluProArgArgGlu----- 383
Db 4127 GAGCAGGACCTCCCGCGCAGAGATTCCCGCACCACCTCCACCCCAATACGCCCCAG 4186
Qy 384 -----AsnLysSerArgArgArgThrThrAsnAspSerAlaAlaSerGluSerProPro 401
Db 4187 CTGTTGGCGCAAGACGACCTCTAATAACAATACGTCCTCATACCTCCCTAATACGCCG 4246
Qy 402 ProllysArgLeuLysThrAsn-----SerTyGlyGlyLysAsp 414
Db 4247 CCCGTCGCCCTGAGCAGCAGTTCGTCGCCCGCCGCGCAGTCCCGGGGCGAGCAAG 4306
Qy 415 ArgGlyGluAspGluSerArgGluArgMetAlaSerGluValThrAsnAsnLysGly 434
Db 4307 AACGGCGAC-----CAAGGCAGC 4324
Qy 435 AsnLeuGluAspArgCysLeuSerCys-----GlyLysLysAsn 447
Db 4325 AACGGCCGACCACTGCGACGAAGTCTTCACTCAAAACCACTCCGCTGTGGCGTCGGAAC 4384
Qy 448 ProValSerPheHisProLeuPheGluGly-----GlyLeuCysGlnSerCysArgAspArg 466
Db 4385 CCA-----GAGGCGCAGCCACTGTGCAATGCCTGC---GGGTTG 4420
Qy 467 PheLeuGluLeu 470
```

Db 4421 TTTTGAATTG 4432

Search completed: November 25, 2002, 02:56:29  
Job time : 4354.34 secs



Mon Nov 25 08:18:36 2002

us-09-

```
13 161 3.3 7791 4 US-08-984-709A-23 Sequence 23, Appl
14 161 3.3 7791 4 US-08-450-272-23 Sequence 23, Appl
15 161 3.3 7808 3 US-08-949-386-22 Sequence 22, Appl
16 161 3.3 7808 3 US-08-450-556-22 Sequence 22, Appl
17 161 3.3 7808 4 US-08-984-709A-22 Sequence 22, Appl
18 161 3.3 7808 4 US-08-450-272-22 Sequence 5, Appl
19 159 3.2 2277 1 US-08-676-967-5 Sequence 5, Appl
20 159 3.2 2277 1 US-08-676-974-5 Sequence 5, Appl
21 159 3.2 2277 2 US-08-098-487-5 Sequence 1, Appl
22 159 3.2 2277 2 US-07-945-283-1 Sequence 1, Appl
23 158.5 3.2 2680 1 US-08-053-035-1 Sequence 93, Appl
24 158.5 3.2 9551 1 US-08-056-200-93 Sequence 93, Appl
25 158.5 3.2 9551 2 US-08-800-644-93 Sequence 3, Appl
26 157 3.2 20235 1 US-07-642-734C-3 Sequence 3, Appl
27 157 3.2 20235 3 US-08-419-009A-3 Sequence 22, Appl
28 156.5 3.2 1736 4 US-09-162-524-2 Sequence 3, Appl
29 156.5 3.2 7808 2 US-08-199-097D-22 Sequence 22, Appl
155 3.1 1681 4 US-09-434-288-7 Sequence 7, Appl
155 3.1 5468 4 US-09-535-008-66 Sequence 66, Appl
155 3.1 5471 4 US-09-535-008-1 Sequence 1, Appl
155 3.1 5477 4 US-09-535-008-74 Sequence 74, Appl
155 3.1 5480 4 US-09-535-008-70 Sequence 70, Appl
155 3.1 5584 4 US-09-535-008-68 Sequence 68, Appl
155 3.1 5587 4 US-09-535-008-64 Sequence 64, Appl
155 3.1 5573 4 US-09-535-008-76 Sequence 76, Appl
155 3.1 5576 4 US-09-535-008-72 Sequence 72, Appl
155 3.1 13842 4 US-09-105-537-30 Sequence 30, Appl
155 3.1 36778 4 US-09-105-537-5 Sequence 5, Appl
155 3.1 38506 3 US-09-320-878-19 Sequence 19, Appl
154 3.1 5467 2 US-07-745-206A-12 Sequence 12, Appl
154 3.1 5467 2 US-08-311-363-12 Sequence 12, Appl
154 3.1 11907 4 US-08-061-376-4 Sequence 4, Appl
154 3.1 14255 1 US-08-320-559-1 Sequence 1, Appl
```

## ALIGNMENTS

```
RESULT 1
US-09-276-531-47
Sequence 47, Application /
Patent No. 6183968
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Hiltman, Jennifer L.
APPLICANT: Yue, Henry
APPLICANT: Reddy, Roopa
APPLICANT: Guegler, Karl J.
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING
NUMBER OF SEQUENCES: 134
RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/276,531
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/079,677
FILING DATE: March 27, 1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
```

```
NAME: Lynn E. Murry, Ph.D.
REGISTRATION NUMBER: 42,918
REFERENCE/DOCKET NUMBER: PA-0008 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 2077 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: TESTU02
CLONE: 1271435
US-09-276-531-47
Alignment Scores:
Pred. No.: 2,1e-46 Length: 2077
Score: 748.00 Matches: 151
Percent Similarity: 81.95% Conservative: 17
Best Local Similarity: 73.66% Mismatches: 36
Query Match: 15.16% Indels: 4
DB: Gaps: 1
US-09-720-086-7 (1-912) x US-09-276-531-47 (1-2077)
QY 707 GlySerProCysAsnAspLeuSerIleValAsnProAlaArgIleGlyLeuTyrGluGly 726
3 GCGCGGCAATGCAAGCATCTTNNAAAT-GTGAATCA-CCGAGAAAGCGCTGATACGCGAC 60
DB 61 AGG---CGGCTCTTCTCGAATTTC-CACCTGCTGAATTACTACGACCCCAAGAGGCT 116
QY 747 AspAspArgProPhePheTyrPheGluAsnValValAlaMetGlyValSerAspLys 766
117 GATGACCGGCGCTCTCTGAGATGTTGAGAAATGTTGACNCTCGGCCGATTCGCGACA 176
DB 767 ArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLysGluValSer 786
177 GCGGACATCTCAAGGTTCCGAGGTGATCCAGTATGATGATGATCCATCAAGTTCT 236
QY 787 AlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMetAsnArgProLeuAla 806
237 GCTGCTCACAGGCCGATATCTTCTGGGCAACCTTACCCGGGATGAACAGGCCGATGA 296
DB 807 SerThrValAsnAspLysLeuGluGlnGlyCysLeuGluHisGlyArgIleAlaLys 826
297 GCATCAAGAAATGATTAATCTGCTGCGAGCTGCTGGAATACATCAATAGGATACCAAG 356
QY 827 PheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGlyLysAspGln 846
357 TTAAGAAAGTACACAAATTAACCAACCAAGTCAATCGATCAACAGGAGAAAAACCA 416
DB 847 HisPheProValPheMetAsnGluLysGluAspIleLeuTyrCysThrGluMetGluArg 866
417 CTTTCCCTGTTGTCATGATGATGCAAGAGATGTTGTTGTTGTTGTTGTTGTTGTTG 476
QY 867 ValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeuAlaArgGln 886
477 NTCTTGGCTTCTGTCGACTACACAGACCTGTCACAACTGGGCGGTGGCGGCGAC 536
DB 887 ArgLeuLeuGlyArgSerTyrSerValProValIleArgHisLeuPheAlaProLeuLys 906
537 AAGCTGCTGGAAAGTCTCGAGCGCTGCTGTCATCCGACCTTTCGCCCTTGAG 596
QY 907 GluTyrPheAlaCys 911
DB 597 GACTACTTGCAATGT 611
RESULT 2
US-09-105-537-1/C
```

**THIS PAGE BLANK (USPTO)**

AF068625 4192 bp mRNA linear ROD 06-DEC-1999  
 LOCUS Mus musculus DNA cytosine-5 methyltransferase 3A (Dnmt3a) mRNA,  
 DEFINITION complete cds.  
 ACCESSION AF068625.2 GI:6449467  
 VERSION AF068625.2  
 KEYWORDS Mus musculus.  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 4192)  
 AUTHORS Okano, M., Xie, S., and Li, E.  
 TITLE Cloning and characterization of a family of novel mammalian DNA  
 (cytosine-5) methyltransferases  
 JOURNAL Nat. Genet. 19 (3), 219-220 (1998)  
 PUBMED 98324766  
 MEDLINE 9662389  
 REFERENCE 2 (bases 1 to 4192)  
 AUTHORS Xie, S., Okano, M., and Li, E.  
 TITLE Direct Substitution  
 JOURNAL Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,  
 Charlestown, MA 02129, USA  
 REFERENCE 3 (bases 1 to 4192)  
 AUTHORS Okano, M., Chijiwa, T., Sasaki, H., and Li, E.  
 TITLE Direct Substitution  
 JOURNAL Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,  
 Charlestown, MA 02129, USA  
 REMARK Sequence update by submitter  
 COMMENT On Nov 18, 1999 this sequence version replaced gi:3327977.  
 FEATURES  
 source  
 1..4192  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="12"  
 /map="4.0 CM"  
 1..4192  
 /gene="Dnmt3a"  
 /gene="Dnmt3a"  
 1..4192  
 /note="Similar to EST sequences deposited in Genbank  
 Accession Numbers AA052791, AA111043, AA154890, AA240794,  
 AA756653, W58898, W59299, W91664, and W91665"  
 217..2943  
 /gene="Dnmt3a"  
 /function="de novo DNA methylation"  
 /note="contains Cys-rich region"  
 /codon\_start=1  
 /product="DNA cytosine-5 methyltransferase 3A"  
 /protein\_id="AAC40177.2"  
 /translation="MPSGPDSTSSSLERDRKGEQEEENRGRKEQESQPSATRK  
 VPRPKRKHPRVESDTPKDPVATTKSPMAODSGPDLNGLDEKSEQPEGS  
 PAAGQGGAPAGBGTETPPRAVENECVTKEGRGASAGEKQKOTNISMKE  
 GSRGLRGLGWESSLRQPMRLTQADDPYISRKRDENLARKAEKAKYIA  
 VNAAVENQASGESQKVEASPAVQPTDPAAPYATTPPEPGAGKMTAKAD  
 BEVEDRGFGIGELVMGLRGSWMPGIVSMVMGSRRAEGLTMMWMMFGDKRSV  
 VCVEKMLPSFCSAFHATYKQPMYRAIYEVLOVASSRAKLPACHDSBDSG  
 KAVEVONKMIEMALGGPQSPGKLEPPEEKNGYETIDMVAPEAAVAPPA  
 MCQNKCFLECAVYQYDDGOSTCTCCGREVLMGNNCCRCCECVDLVPG  
 KAPKSTTEPKYKEITIDERTREPLRYEVROKRNIEDICISGLNLTLEHLFTGG  
 AAQAAIKEDPMNMGCHGTGLLRREDWPSRLQMFANNDQSPDPKVPYPA  
 EKRPILRVSLFDGIALGLVLDLQVDRYLAISEVCESDITVGNRRHGKIMYGD  
 VRSVQKHIOEMGPDLVIGSPCNDSIVNPAKGLYEGTGRLFEFVRLDAPK  
 EGDRPFFMLFENVVMGVSDKRDISFLSNPVMIDAKESAAHRAHYFWGLDAPK  
 RPLASTVNDKLEOECLHGRIRAKFSVRTITRSNIIKQKDHFEVMEKEDILM  
 CTMERVGFPHVTVDVSNMSRLAKRLLRSMSVPIRLHPLKKEYFACV"

Alignment Scores:  
 Pred. No.:  
 Score:

3.25e-211  
 4909.00

Length:  
 Matches:  
 4192  
 908

Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 100.00%  
 DB: 10  
 Gaps: 0

US-09-720-086-5 (1-908) x AF068625 (1-4192)

QY 1 MetProSerSerGlyProGlyAspThrSerSerSerLeuGluAlaGluAspAspArg 20  
 Db 217 ATGCCCTCCAGCGGGCCCGGAGACACACACACTCTCTCTGAGCCGAGATATATCA 276  
 QY 21 LysGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 40  
 Db 277 AAGGAAG 336  
 QY 41 ThrAlaArgLysValGlyArgProGlyArgLysValGlyArgLysValGlyArg 60  
 Db 337 ACGCCCGAGAGGTGGAGGCTTGGCCCGAAGCGCACACCCAGCGGTGGAACCACT 396  
 QY 61 AspThrProLysAspProAlaValThrThrLysSerGluProMetAlaGluAspSerGly 80  
 Db 397 GACACCCCAAGAGACCCAGAGAGACCACTTCAAGCTTCAAGCTTCAAGCTTCAAG 456  
 QY 81 ProSerAspLeuLeuProAsnGlyAspLeuGluLysArgSerGluProGluProGlu 100  
 Db 457 CCTCAGATCTGCTCAACCAATGAGACTTGAGAGAGCGAGTGAACCCCAACCTGAGAG 516  
 QY 101 GlySerProAlaAlaGlyGluLysGlyGlyAlaProAlaGluGluGluGluGluThr 120  
 Db 517 GGGAGCCAGCTGCGAGGCGAGAAAGGTGGGCCCCCTGAGAGAGAGAGAGAGAGAG 576  
 QY 121 ProProGluAlaSerArgAlaValGluAsnGlyLysValThrLysGluGlyArgGly 140  
 Db 577 CCACAGAGAGCTCCAGAGCTGTGAGAGATGCTGCTGTGAGAGAGAGAGAGAGAGAG 636  
 QY 141 AlaSerAlaGlyGluGlyLysGluGluGluGluGluGluGluGluGluGluGluGlu 160  
 Db 637 GCTTGTGAG 696  
 QY 161 GlySerArgGlyArgLeuArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 180  
 Db 697 GCTTCCCGGGGGCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 756  
 QY 181 MetProArgLeuThrPheGluAlaGlyAspProGlyGlyGlyGlyGlyGlyGlyGly 200  
 Db 757 ATGCCAAGACTCACCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 816  
 QY 201 GluThrLeuAlaArgThrLysArgGluAlaGluLysLysAlaLysValIleAlaValMet 220  
 Db 817 GAGTGGCTGACAGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 876  
 QY 221 AsnAlaValGluGluAsnGluAlaSerGlyGlySerGluLysValGluGluAlaSerPro 240  
 Db 877 AATGCTGTGGAAGAAGACAGGCTTGGAGAGTCTCAGAGAGTGAAGAGAGAGAGAG 936  
 QY 241 ProAlaValGluGluProThrAspProAlaSerProThrValAlaThrThrProGluPro 260  
 Db 937 CTGCTGTGAG 996  
 QY 261 ValGluGlyAspAlaGlyAspLysAsnAlaThrLysAlaAlaAspAspGluProGluTyr 280  
 Db 997 GTAG 1056  
 QY 281 GluAspLysArgLysPheGlyIleGlyGluLeuValIleProLysLysLeuArgGlyPheSer 300  
 Db 1057 GAGATGCGCGGGCTTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1116  
 QY 301 TrpThrProGlyArgGlyIleValSerTrpThrMetThrGlyArgSerArgAlaGluGly 320  
 Db 1117 TGTGTGCGAG 1176  
 QY 321 ThrArgThrValMetThrPheGlyAspGlyLysPheSerValValLysValGluLysLeu 340

**THIS PAGE BLANK (USPTO)**

1177 ACTGCTGATCATGTGTTGGAGATGCAAGTTCCTCAGTGTGTGTGGAGAGCTC 1236  
 QY 341 MetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnProMet 360  
 Db 1237 ATGGCGCTGACCTCTCTGTGAGTGATTTCCACCGAGCCACTTCAACACAGAGCCCATG 1296  
 QY 361 TyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPhe 380  
 Db 1297 TACCGCAAGCCATCTAGAAAGTCTCCAGGTGGCCAGACCCGCGGGAAGCTGTTT 1356  
 QY 381 ProAlaCysHisAspSerAspGluSerAspSerGlyLysAlaValAlaGlnAsnLys 400  
 Db 1357 CCACCTTGCCATGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1416  
 QY 401 GlnMetIleGluTTPAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGluPro 420  
 Db 1417 CAGATGATGAAATGGCCCTCGGTGGCTTCCAGCCCTCGGTGGCTTCCAGCCCTGAGCCA 1476  
 QY ProGluGluGluLysAsnProTyrLysGluValTyrThrAspMetTPValGluProGlu 440  
 Db 1477 CCAGAGAAGAGAAATCTTACAGAGAGATTTACACGACATGTGGGTGGAGCTGAA 1536  
 QY 441 AlaAlaAlaTyrAlaProProProProAlaLysLysProAlaGlySerThrThrGluLys 460  
 Db 1537 GCAGCTGCTTACGCCCAACCCCAACCAAGAAACCCCAAGAGACAAACAGAGAAA 1596  
 QY 461 ProLysValLysGluLysIleAspGluArgThrArgGluArgLysValTyrGluValArg 480  
 Db 1597 CCTAAGCTCAAGAGATCATGTATGAGCGCAAGAGAGCGGTGTGTATGAGGTGGC 1656  
 QY 481 GlnLysCysArgAsnIleGluAspLysIleCysArgLysSerLeuAsnValThrLeu 500  
 Db 1657 CAGAGTGCAAAACATCGAGACATTTGTATCTCATGTGGAGCTCAAGTCAACCTTG 1716  
 QY 501 GlnLysProLeuPheIleGlyGlyMetCysGlnAsnCysLysAsnGlyPheLeuGluCys 520  
 Db 1717 GAGCACCACTCTTATGAGGAGCATGTGCCAGAACTTAAGAACTGTTCTTGGAGTGT 1776  
 QY 521 AlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyValArg 540  
 Db 1777 GCTTACAGATGAGCAGATGGTACAGTCCCTATGACCATCTGCTGTGGGGGGGT 1836  
 QY 541 GluValLeuMetCysGlyLysAsnAsnAsnCysArgCysPheCysValGluCysValAsp 560  
 Db 1837 GAAGTCTCATGTGTGGAAACAACTGTGACAGTCTTTGTGTGAGTGTGTGAT 1896  
 QY LeuLeuValGlyProGlyValAlaAlaGlnAlaAlaIleLysGluAspProTyrAsnCysTyr 580  
 Db 1897 CTCTTGTGGGGGCCAGAGCTGTCTAGGCAAGCATTTAAGAAAGCCCTGGAACTGTAC 1956  
 QY 581 MetCysGlyHisLysGlyThrTyrGlyLeuLeuArgArgGluAspTyrProSerArg 600  
 Db 1957 ATGTGCGGCAATAAGGCACTATGGCTGTGCGAAGACGGAAAGACTGGCTTCTCGA 2016  
 QY 601 LeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyrPro 620  
 Db 2017 CTCGAGATGTTCTTTGCCAATAACCATGACCAAGAAATTGACCCCCCAAGGTTTACCCA 2076  
 QY 621 ProValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIleAla 640  
 Db 2077 CTTGTGCGAGCTGAGAAAGAGAGCCCATCCGCGTGTCTCTTTTGATGGGATGCT 2136  
 QY 641 ThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSerGlu 660  
 Db 2137 ACAAGGCTCTGTGTGTAAGGACTGTGGCATCCAAATGAGACCGGTATATGCTCCGAG 2196  
 QY 661 ValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyrVal 680  
 Db 2197 GTGTGTGAGACTCCATCAGCGTGGCATGTGTGCGGACCAAGGAAAGATCATGTACGTC 2256  
 QY 681 GlyAspValArgSerValThrGlnLysHisIleGlnGluTTPGlyProPheAspLeuVal 700  
 Db 2257 GGGAGCTGCGAGCTCACACAGAGCATATCCAGAGTGGGGCCCATTCGACTGGGTG 2316

QY 701 IleGlyLysSerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeuTyr 720  
 Db 2217 ATGGAGGAGAGTCCCTGGCAATGACTCTTCATTTGTCAACCTCGCCCAAGGAGCTTTAT 2376  
 QY 721 GluGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgProLys 740  
 Db 2377 GAGGATCTGGCCGCTCTTCTTTGAAGTTTACCGCTCTCTGCATGATGGCGGCCAAG 2436  
 QY 741 GluGlyAspAspArgProPhePheTyrPheGluLeuAsnValAlaAlaMetGlyValSer 760  
 Db 2437 GAGGAGATGATCGCCCTCTTCTTGTGCTCTTGTGAGAAATGGTGGCCATGGGGCTTAGT 2496  
 QY 761 AspLysArgAspIleSerArgPheLeuGluLysAsnProValMetIleAspAlaLysGlu 780  
 Db 2497 GACAAAGGAGACATCTCGCATTTCTTGAGCTTAAACCCGTGATGTAGACGCAAGAA 2556  
 QY 781 ValSerAlaAlaHisArgValArgTyrPheTyrGlyAsnLeuProGlyMetAsnArgPro 800  
 Db 2557 GTGTCTGTCACACAGGAGCCGCTTACTTCTGGGGTAACTTCTGTCATGAAAGAGCT 2616  
 QY 801 LeuAlaSerThrValAsnAspLysLeuGluLeuGlnGlnCysLeuGlnHisGlyArgIle 820  
 Db 2617 TTGGCATCCACTGTGAATGATTAAGCTGAGCTGCAAGAGTGTCTGGAGCAGCGCAATA 2676  
 QY 821 AlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGlyLys 840  
 Db 2677 GCCAAGTTCAAGCAAGTGAAGAGACCATTTACACAGCTCAAACTTATTAAGCAGGGCAA 2736  
 QY 841 AspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTyrCysThrGluMet 860  
 Db 2737 GACCAGATTTCCCGCTCTTCAATGAACGAAAGAGAGACATCTGTGTGTCATCGAAATG 2796  
 QY 861 GluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeuAla 880  
 Db 2797 GAAAGGATGTTGGCTTCCCGCTCACTACACAGACTCTCCAAATGAGACCGGCTTGGCG 2856  
 QY 881 ArgGlnArgLeuLeuGlyArgSerTyrSerValProValIleArgHisLeuPheAlaPro 900  
 Db 2857 AGCAGAGAGACTGTGGGCGCATCTGTGAGCGGTGCGGTATCCGCCACCTTGTGCTCCG 2916  
 QY 901 LeuLysGluTyrPheAlaCysValIle 908  
 Db 2917 CTGAAGGAATTTTGTGTTGTG 2940

RESULT 2

**THIS PAGE BLANK (USPTO)**

RESUL  
AAT21884  
ID AAT21884 standard; cDNA to mRNA; 301 BP.  
XX  
AC AAT21884;  
XX  
DT 14-AUG-1996 (first entry)

XX  
DE Human gene signature HUMGS03426.  
XX  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9514772-A1.  
XX  
PD 01-JUN-1995.  
XX  
PF 11-NOV-1994; 94WO-JP01916.  
XX  
PR 12-NOV-1993; 93JP-0355504.  
XX  
PA (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
XX  
PI Matsubara K, Okubo K;  
XX  
DR WPI; 1995-206931/27.  
XX  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
XX  
PS Claim 1; Page 991-992; 2245pp; Japanese.  
XX  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in AAT19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
XX  
SQ Sequence 301 BP; 96 A; 65 C; 63 G; 66 T; 11 other;  
Query Match 6.6%; Score 285.2; DB 16; Length 301;  
Best Local Similarity 95.3%; Pred. No. 1.5e-48;  
Matches 287; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
QY 3519 GATCAGATAGGAGCACAAGCAGGGGACGGAAGAGAGAGACACTCAGGCGGCAGCATTCC 3578  
Db 1 GATCAGATAGGAGCACAAGCAGGGGACGGAAGAGAGAGACACTCANGCGGCAGCATTCC 60  
QY 3579 CTCCCAGCCACTGAGCTGTCTGTCGACCACTTCCTGGTCACGCAAAACAGAACCCAGT 3638  
Db 61 CTNCCAGCCACTGAGCTGNCGTGCCAGCACCATTCTCTGGTCACGCAAAACAGAACCCATT 120  
QY 3639 TAGCAGCAGGGAGACGAGAACACCACACAAGACATTTTCTACAGTATTTTCAGGTGCCTA 3698  
Db 121 TAGCAGCAGGGAGACGAGAACACCACACAAGACATTTTCTACAGTATTTTCAGGTGCCTA 180  
QY 3699 CCACACAGGAACCTTGAAGAAAATCAGTTTCTAGAAGCCGCTGTTACCTCTTGTTTACA 3758  
Db 181 CCACACAGGAACCTTGAAGAAAATCAGTTTCTAGGNNNGCTGTTACCTCTTGTTTACA 240  
QY 3759 GTTTATATATATATGATAGATATGAGATATATATATAAAAGGTACTGTTAACTACTGTAC 3818  
Db 241 GTTTATATATATATGATAGATATGNGNTATATATATAAAAGGTACTGTTAACTACTGTAA 300

QY 3819 A 3819  
Db 301 A 301

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 22, 2002, 03:58:32 ; Search time 5173.7 Seconds  
(without alignments)  
5107.633 Million cell updates/sec

Title: US-09-720-086-5  
Perfect score: 4903  
Sequence: 1 MPSSGGPGDTSSSLEREDDR.....MSVPYIRHLPAPIKEYFACV 908

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xld  
-O/cgpn2.1/USPTO.spool/US9720086/runat.18112002.092832.22195/app.query.fasta\_1.4252  
-DB=GenEmbl -QPM=fastap -SUFRTX=rge -MINMATCH=0 -LDOFCU=0 -LDOFEXT=0  
-UNITS-bits -START=1 -END=1 -MATRIX=blomus62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09720086 @CGN 1.1 11308 @runat.18112002.092832.22195 -NCPU=6 -ICPU=3  
-NO\_XLPEXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMECUT=120  
-WEN\_TIMECUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	4909	100.0	4192	10 AF068625	AF068625 Mus muscu
2	4903	99.9	4094	10 BC007466	BC007466 Mus muscu
3	4729.5	96.3	3005	9 AF067972	AF067972 Homo sapi
4	4729.5	96.3	4258	9 AF331856	AF331856 Homo sapi
5	2571.5	52.4	123936	2 AC120824	AC120824 Rattus no
6	2412	49.1	2191	9 AK025230	AK025230 Homo sapi
7	2221.5	45.3	168651	9 AC009474	AC009474 Homo sapi
8	2211	45.0	4135	10 AF068627	AF068627 Mus muscu
9	2211	45.0	4278	10 AF151970	AF151970 Mus muscu
10	2206	44.9	4195	10 AF068626	AF068626 Mus muscu
11	2206	44.9	4338	10 AF151969	AF151969 Mus muscu
12	2203.5	44.9	4163	10 AF151974	AF151974 Mus muscu
13	2200.5	44.8	176897	2 AC112586	AC112586 Rattus no
14	2200	44.8	4335	9 AF331857	AF331857 Homo sapi
15	2198.5	44.7	4223	10 AF151973	AF151973 Mus muscu
16	2195.5	44.7	4267	9 AF176228	AF176228 Homo sapi
17	2187	44.6	4145	9 AF156488	AF156488 Homo sapi
18	1938.5	39.5	3946	10 AF068628	AF068628 Mus muscu
19	1938.5	39.5	4089	10 AF151971	AF151971 Mus muscu
20	1933.5	39.4	4006	10 AF078427	AF078427 Mus muscu
21	1933.5	39.4	4149	10 AF151972	AF151972 Mus muscu
22	1931	39.3	3974	10 AF151976	AF151976 Mus muscu
23	1926	39.2	4034	10 AF151975	AF151975 Mus muscu
24	1908.5	38.9	3897	9 AF156487	AF156487 Homo sapi
25	1835.5	37.4	204534	2 AC118195	AC118195 Mus muscu
26	1835.5	37.4	208874	2 AC111092	AC111092 Mus muscu
27	1763.5	35.9	2848	5 AF135438	AF135438 Dario rer
28	1748.5	35.6	197652	9 AC012074	AC012074 Homo sapi
29	1623.5	33.1	3017	9 AK001191	AK001191 Homo sapi
30	1504	30.6	1758	9 BC018214	BC018214 Homo sapi
31	1445.5	29.4	210269	2 AC116459	AC116459 Mus muscu
32	1184	24.1	185413	2 AC119009	AC119009 Rattus no
33	1032.5	21.0	119630	2 AC112040	AC112040 Rattus no
34	996.5	20.3	167568	2 AC111734	AC111734 Rattus no
35	994	20.2	2127	9 AF129257	AF129257 Homo sapi
36	872.5	17.8	2057	9 AF129268	AF129268 Homo sapi
37	785	16.0	2008	9 AF129269	AF129269 Homo sapi
38	748	15.2	2077	6 AF129189	AF129189 Sequence
39	728.5	14.8	119630	2 AC112040	AC112040 Rattus no
40	692.5	14.1	1397	9 BC002560	BC002560 Homo sapi
41	688	14.0	1705	9 AF220524	AF220524 Homo sapi
42	644.5	13.1	1589	10 AF220524	AF220524 Mus muscu
43	644.5	13.1	1645	10 MM040467	MM040467 Mus muscu
44	561.5	11.4	244329	2 AC107644	AC107644 Mus muscu
45	490.5	10.0	204534	2 AC118195	AC118195 Mus muscu

ALIGNMENTS

RESULT 1

AF068625 4192 bp mRNA linear ROD 06-DEC-1999  
LOCUS Mus musculus DNA cytosine-5 methyltransferase 3A (Dnmt3a) mRNA,  
DEFINITION complete cds.  
ACCESSION AF068625 GI:6449467  
VERSION AF068625.2  
KEYWORDS  
SOURCE  
ORGANISM Mus musculus.  
REFERENCE  
AUTHORS Okano, M., Xie, S. and Li, E.  
TITLE Cloning and characterization of a family of novel mammalian DNA  
(cytosine-5) methyltransferases  
JOURNAL Nat. Genet. 19 (3), 219-220 (1998)  
MEDLINE 98324766  
PUBMED 9662389  
REFERENCE  
AUTHORS Xie, S., Okano, M. and Li, E.  
TITLE Direct Submission  
JOURNAL Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA  
REFERENCE  
AUTHORS Okano, M., Chijiwa, T., Sasaki, H. and Li, E.  
TITLE Direct Submission  
JOURNAL Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA  
REMARK  
COMMENT Sequence update by submitter  
FEATURES  
Location/Qualifiers  
1..4192  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="12"  
/map="4.0 cM"  
1..4192  
/gene="Dnmt3a"  
1..4192  
/gene="Dnmt3a"  
/note="similar to EST sequences deposited in GenBank  
Accession Numbers AA052791, AA11043, AA154890, AA240794,  
AA756653, W58898, W59299, W91664, and W91665"  
217..2943  
/gene="Dnmt3a"  
/function="de novo DNA methylation"  
/note="contains Cys-rich region"  
/codon\_start=1  
/product="DNA cytosine-5 methyltransferase 3A"  
/protein\_id="AAC40177.2"  
/db\_xref="GI:6449468"  
/translation="MPSSGPGDTSSTSLEREDDRKEGEQENRKEERQEPSATARK  
VGRPKRKHFPVSSDTTPKPAVTTKSPMAQDSGSDLLPNGDLKRSRQPEGS  
PAAGKRGAPAGEGGTETTPESRAVENCCVTEKRGASAGEKQKOTNIESKME  
GSRRLRGGLGWESLKRQEMPRLTIFQAGDPYIISKRGDMLARWKREAEKAVIA  
VMNAVEENQASGESLKRQEMPRLTIFQAGDPYIISKRGDMLARWKREAEKAVIA  
VEYEDGGRFGIBLVKGLRQFSPWPRGLVSNWMTGRSRAAGTRVWVWFGDKFSV  
KVEKLMPLSSFCFAHQATYINKQPMYKAIYEVLQVASSRAGKJFFACHDSDESDSG  
KAVEVQNMIEWALGFGQFGPKGLEPPPEEKNPYKEVITDMMVEPEAAAYAPPPA  
KKPKSTTKPKPKVEIIDETRELRYVEVRQKRNIEDICISCSGLNVTLEHLLVFG  
MCQNCNCFLECAQYDDQYQSYCTCCGQREVLMCGNNCCRCFCVCEVDLLVFG  
AAQAIKEDPMNVCYMGKGTGYLLRRREDWPSRLQMFANNHDSFDPKVPVPPVA  
EKRRITRLVSLFDGTLGLDGLQVDRYIASEVEDSITVGMVHKGKIMVYGD  
VRSVTQKHQIOWGPFVIGGSPNDLSIVNPAKGLYEGTGRFLFFEFYRLLHDAPK  
EGDDPFFWLFENVAVGVSDKRDISRFLESNVMDIAKEVSAHAHRARYFWGLPDMN  
RPLASTVNDKLELECELEHGRITAKFSKVTITTRNSIKQGDQHFVFPVFNKEEDILW  
CTEMERVFGFPVHYTDVSNMSRLARQLLRGSWSVPIRHLFAPLKFYFACV"  
BASE COUNT 1094 a 1072 c 1186 g 840 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3,25e-211 Length: 4192  
Score: 4909.00 Matches: 908

Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 10 Gaps: 0  
US-09-720-086-5 (1-908) x AF068625 (1-4192)  
QY 1 MetProSerSerGlyProGlyAspThrSerSerSerSerLeuGluArgGluAspArg 20  
DB 217 ATGCGCTCCAGCGCGCGCGGACACACAGCAGCTCTCTCTGAGCGGAGGATGATCGA 276  
QY 21 LysGluGlyGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 40  
DB 277 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 336  
QY 41 ThrAlaArgLysValGlyArgProGlyArgProGlyArgLysArgLysArgLysArgLysArg 60  
DB 337 ACGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 396  
QY 61 AspThrProLysAspProAlaValThrThrLysSerGlnProMetAlaGlnAspSerGly 80  
DB 397 GACACCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 456  
QY 81 ProSerAspLeuLeuProAsnGlyAspLeuGluLysArgSerGluProGlnProGluGlu 100  
DB 457 CCCTCAGATCTGTACCTACCAATGGAGACTTGGAGAGCGGAGTGAACCCCACTCAGGAG 516  
QY 101 GlySerProAlaAlaGlyGlnLysGlyGlyAlaProAlaGluGlyGluGlyThrGluThr 120  
DB 517 GGGAGCGCGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 576  
QY 121 ProProGluAlaSerArgAlaValGluAsnGlyCysCysValThrLysGluGlyArgGly 140  
DB 577 CCACCGAGAGCTCCAGAGCTGTGGAGAAATGGCTGTGTGTGACCAAGAGGCGCTGGA 636  
QY 141 AlaSerAlaGlyGluGlyLysGluGlnLysGlnThrAsnIleGluSerMetLysMetGlu 160  
DB 637 GCCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 696  
QY 161 GlySerArgGlyArgLeuArgGlyGlyLeuGlyThrGluSerSerLeuArgGlnArgPro 180  
DB 697 GGCTCCCGGGCGGAGTGGAGAGTGGCTGGAGTGGCTGGAGTGGCTGGAGTGGCTGGAGTGGCTGGAG 756  
QY 181 MetProArgLeuThrPheGlnAlaGlyAspProTyrTyrIleSerLysArgLysArgAsp 200  
DB 757 ATGCCAAGACTCACCTTCCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 816  
QY 201 GluThrLeuAlaArgThrLysArgGluAlaGluLysLysAlaLysValIleAlaValMet 220  
DB 817 GAGTGGCTGGACACCTTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 876  
QY 221 AsnAlaValGluGluAsnGlnAlaSerGlyGluSerGlnLysValGluGluAlaSerPro 240  
DB 877 AATGCTGTGGAAGAGAACACAGGCGCTCTGAGAGAGTCTCAGAGAGTCTGAGAGGAGGAGGAGGAG 936  
QY 241 ProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrThrProGluPro 260  
DB 937 CTGTGTGTGCAGCAGCGCCACGAGACCTGTCTCTCCGACTGTGGCCACACCCCTCAGACCA 996  
QY 261 ValGlyGlyAspAlaGlyAspLysAsnAlaThrLysAlaAlaAspAspGluProGluTyr 280  
DB 997 GTAGAGGAGGAGTGTGGGAGCAGAGTGTACCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1056  
QY 281 GluAspGlyArgGlyPheGlyIleGlyGluLeuValThrGlyLysLeuArgGlyPheSer 300  
DB 1057 GAGGATGGCGGGCGCTTGGCATTTGAGAGCTGTGTGGGGGAACTTCTCGGGGCTTCTCC 1116  
QY 301 TrpThrProGlyArgIleValSerThrThrThrGlyArgSerArgAlaAlaGlyGly 320  
DB 1117 TGTGTGCGCAGCGCGAAATGT 1176  
QY 321 ThrArgThrThrValMetThrPheGlyAspGlyLysPheSerValValCysValGluLysLeu 340

Db 1177 ACTCGTGGGTCATGTCGTCGAGATGGCAATTTCTCACTGCTGTGTGGAAGGCTC 1236  
 Qy 341 MetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnIysGlnProMet 360  
 Db 1237 ATGCCCTGAGCTCTCTTCTGACGTGATTCACCAAGGCCACTTAACAAGCAAGGCCCATG 1296  
 Qy 361 TyrArgIlyAlaIleTyrGluValIleGlnValAlaSerSerArgAlaGlyIlyLeuPhe 380  
 Db 1297 TACCGAAGGCCATCTACAGAACTCTCCAGTGGCCAGCAGCGGTGCCGGAAGCTGTTT 1356  
 Qy 381 ProAlaCysHisAspSerAspGluSerAspSerGlyIlyAlaIValGluValGlnAsnIys 400  
 Db 1357 CCAGCTTGCCATACAGTATGAAAGTGAAGTGGCAAGGCTGTGAAGTGCAGAAACAG 1416  
 Qy 401 GlnMetIleGlnTyrAlaLeuGlyIlyPheGlnProSerGlyProIysGlyLeuGlnPro 420  
 Db 1417 CAGATGATTAATGAGGCCCTCGGTGGCTTCAAGCCCTCGGTGCTTAAGGCCCTGGAGCCA 1476  
 Qy 421 ProGluGluGluIlyAsnProTyrIlyGluValIlyThrAspMetTyrValGluProGlu 440  
 Db 1477 CCAGAAAGAGAGAAATCCTTCAAGGAAGTTTACACCAACATGTGGGTGGAGCCTGAA 1536  
 Qy 441 AlaAlaIleTyrAlaProProProProAlaIlyAsnProArgIlySerThrThrGluIys 460  
 Db 1537 GCACTGCTTACGCCCCCAAGCCCAAGCAAGAAACCCAGAAAGACACAGCAAGAGAAA 1596  
 Qy 461 ProIlyValIlyGluIleIleAspGluArgThrArgGluArgLeuValIlyGluValArg 480  
 Db 1597 CCTAAGTCAAGAGATCATTTGATGAGCCGCAAGGAGGCGGTGGTGTATGAGTCCG 1656  
 Qy 481 GlnIlyCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValIlyLeu 500  
 Db 1657 CAGAAAGTGCAGAAATCGAGGACATTTGTATCTCATGTGGAGACCTCAATGCACCTG 1716  
 Qy 501 GlnHisProLeuPheIleGlyIlyMetCysGlnAsnCysIlyAsnIysPheLeuGluCys 520  
 Db 1717 GAGACCCACCTCTTCAATGAGGAGTGTGCCAGAACTGTAAGAACTGCTTCTGGAGTGT 1776  
 Qy 521 AlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyIlyArg 540  
 Db 1777 GCTTACAGATACAGACGATGGGTACCACTCTATTGCACATCTGCTGTGGGGGCGGT 1836  
 Qy 541 GluValIleuMetCysGlyIlyAsnAsnAsnCysCysArgCysPheCysValGluCysValAsp 560  
 Db 1837 GAAGTCTCATGTGTGGGAAACAACACTGCTGAGGTGCTTTGTGTCGAGTGTGTGAT 1896  
 Qy 561 LeuLeuValGlyProGlyIlyAlaIleGlnAlaAlaIleIlyLeuIlyAspProTyrPheCysTyr 580  
 Db 1897 CTCTTGTGGGGCCAGAGAGCTGCTCAGGCGACCATTAAGGAAGACCCCTGGAACCTGCTAC 1956  
 Qy 581 MetCysGlyIlyIleIysGlyThrTyrGlyIlyLeuLeuArgArgGluAspTyrProSerArg 600  
 Db 1957 ATGTGCGGCGATTAAGGCGACCTATGGCTCTCTCGAAAGCGGAAGACTGGCTTCTCGA 2016  
 Qy 601 LeuGlnMetPhePheAlaAsnAsnHisAspGlnGluIlyPheAspProIysValIlyTyrPro 620  
 Db 2017 CTCAGATGTTCTTTGGCATTAACATGACAGAAATTTGACCCCAAGGTTTACCA 2076  
 Qy 621 ProValProIleGluIlyAsnArgIlyProIleArgValIleuSerIleuPheAspGlyIlyAla 640  
 Db 2077 CCGTGTGCACTAGAGAGAGAGGCCATCGGCTGTCTCTCTTGTGTGGATTTGCT 2136  
 Qy 641 ThrGlyIlyLeuValIleuIlyAspLeuGlyIleGlnValAlaAspArgTyrIleAlaSerGlu 660  
 Db 2137 ACAGGGCTCTGTGCTGAAAGGACCTGGCATCCAGTGAACCGCTTCAATGCTCCGAG 2196  
 Qy 661 ValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyIlyIleMetTyrVal 680  
 Db 2197 GTGTGTAGAGATCTCATCAAGTGTGCGATGTGTGGCAGCAAGGAAAGATCATCTACGTC 2256  
 Qy 681 GlyAspValArgSerValIlyThrGlnIlyHisIleGlnIlyIlyProIlyPheAspLeuVal 700  
 Db 2257 GGGGACGTCCGACGCGTACACAGAAAGCATATCCAGAGATGGGGCCCAATTCGACTGGTG 2316

Qy 701 IleGlyIlySerProCysAsnAspLeuSerIleValAsnProAlaArgIlyGlyLeuTyr 720  
 Db 2317 ATTGAGGCGAGTCCCTGCAATGACCTTCATTTGCAACCCGCGCAAGGACATTTAT 2376  
 Qy 721 GluGlyIlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgProIys 740  
 Db 2377 GAGGTAATGCGCGCTCTTCTTTGAGTTCACCCGCTCTCTCAATGATGCGGCGCCAAAG 2436  
 Qy 741 GluGlyAspAspArgProPhePheTyrLeuPheGluAsnValAlaIleMetGlyIlySer 760  
 Db 2437 GAGGAGATGATCGCCCTCTTCTGCTCTTTGAGAAATGTGTGGCCATGGGCGTTAT 2496  
 Qy 761 AspIlyAspAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaIysGlu 780  
 Db 2497 GACAAAGAGGACATCTCGGATTTCTTGATCTAACCCCGTATGATGATGACCCCAAGAA 2556  
 Qy 781 ValSerIleAlaHisArgAlaArgTyrPheTyrPheTyrGlyAsnLeuProGlyMetLeuAspPro 800  
 Db 2557 GTGTGCTGCACACAGGCGCGTTACTTCTGGGTAACTTCTGGCATGAAACAGGCT 2616  
 Qy 801 LeuAlaSerThrValAsnAspIlyLeuGluGlnIlyCysLeuGlnIlyHisGlyArgIle 820  
 Db 2617 TTGGCATTCACCTGTATATTAAGCTGAGCTGCAGAAAGTGTCTGGAGCACCGAGATA 2676  
 Qy 821 AlaIlyPheSerIlyValArgThrIleThrThrArgSerAsnSerIleIysGlnIlyIys 840  
 Db 2677 GCCAAGTCAAGAAAGTGAAGACCATTAACACAGGTCAACTTAAGCAAGGCGCAAA 2736  
 Qy 841 AspGlnHisPheProValPheMetLeuGlyIlyGluAspIleLeuTyrCysThrGluMet 860  
 Db 2737 GACCACATTTTCCCGCTCTTCAAGCAAGAGAGACATCTGTGTGCACTGCAAAAG 2796  
 Qy 861 GluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeuAla 880  
 Db 2797 GAAAGGTGTTTGCTTCCCGTCCACTACACAGAGCTTCCAACTGAGCCGCTTGGGG 2856  
 Qy 881 ArgGlnArgLeuLeuGlyIlyArgSerTyrSerValProValIleArgHisLeuPheAlaPro 900  
 Db 2857 AGGCAAGACCTGTGGCGCATGTGAGACGTGCGGTATCCGCCACTTCCGCTCCG 2916  
 Qy 901 LeuIlyGluTyrPheAlaCysVal 908  
 Db 2917 CTGAAGAAATATTTGCTGTGTG 2940  
 RESULT 2  
 LOCUS BC007466 4094 bp mRNA linear ROD 07-AUG-2002  
 DEFINITION Mus musculus, DNA methyltransferase 3A, clone MGC:5662  
 ACCESSION IMAGE:3492853, mRNA, complete cds.  
 VERSION BC007466  
 KEYWORDS BC007466.1 GI:13938620  
 SOURCE MGC.  
 ORGANISM house mouse.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 4094)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-MAY-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 CONTACT: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The i.M.A.G.E. Consortium (iMNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) mcd@paxil.stanford.edu  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 8 Row: n Column: 11  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6681208.

## FEATURES

source

1. 4094  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /map="C2ECH II"  
 /clone="MGC:5662 IMAGE:3492853"  
 /tissue\_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMV." (clone lib="NCI CGAP\_Lu29")  
 /lab\_host="DH10B"  
 /notes="Vector: pCMV-SPORT6"  
 245 2971

## CDS

/codon\_start=1  
 /product="DNA methyltransferase 3A"  
 /db\_xref="GI:13938621"  
 /db\_xref="GI:13938621"  
 /db\_xref="LocusID:13435"  
 /translation="MPSSGGDTSSSSLREDDREKGEQENKGEQEPATARK VGRQKRKHPPVSSSTDPKDPVATTKSQPMADSDPSDLDLNGDLKXSEPOPEEGS PAAGQKGAPAGEGCTETTPPEASRAVENGCCTVTEGRGASAGEKQKPTNIESMOE GGRGLRGLOMGESLQRPMPLRTFOAGDPYIYISKRKRDEWLARKREAEKAKVIA VNNVAENQAGSOKVEEASPPAQQDTPASPTVATTPPEVGGDAGDNATKAAAD EPEYEDGFGTIGELVWGLKRGFSWMPGRISVWMTGRSRAAEGTRVWVWFGDKFSV VCVEKMLPSGFCGAFHQATYKQPMYKRAIYEVQLVASSRAGLFPACHDSDSG KAVQNKQMEWALGGFPGSPKGLPEPEEKNPYKEVYDMWVEPEAAAYAPPPPA KKPRLSTTEKPKVKEIIDERTRELRVVEVROKRNIEDICISGSLNVTLEHPLFTGG MCQCNKFCLEAYODDGVQSVCTCCGREVLMCGNNCCRCFCVCEVDLLVGPQ AAAQAIKEDPNVYCMGHGTGYGLRRRDEWPSRLQMFANNHQEDPDPKVPVPPA EKRPRLVSLPDGILGLVLDGLQVDRYIASEVCEDSITVGMVHROKLMYVGD VRSVTKHIOEWGPDVLIIGVSPNDLSIVNPAKGLVEGTRFLPFYFLLHLDARPK EGDPRFPWLFDENVVGVSDKRDISRFLSNPVMIDAKEVSAHRAHYFPMGLPCMN RPLASTVNDKLEOECLEHGRIRKIDRFTITRNSIKQKDOHFPFVPMNEKDLW CTMERVFGFPVHYTDVSNMSRLARQLLGRSWSVPVIRHLFAPLKEYFACY"

BASE COUNT 1067 a 1057 c 1164 g 806 t

## ORIGIN

Alignment Scores:  
 Pred. No.: 5.9e-211 Length: 4094  
 Score: 4903.00 Matches: 907  
 Percent Similarity: 99.89% Conservativeness: 0  
 Best Local Similarity: 99.89% Mismatches: 1  
 Query Match: 99.88% Indels: 0  
 DB: 10 Gaps: 0

US-09-720-086-5 (1-908) x BC007466 (1-4094)

Qy 1 MetProSerGlyProGlyAspThrSerSerSerLeuGluArgGluAspArg 20  
 Db 245 ATGCCCTCCAGCGCGCGCGGACACCGAGCTCTCTCTGAGCGGAGGATATCGA 304  
 Qy 21 LysGluGlyGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 40  
 Db 305 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 364  
 Qy 41 ThrAlaArgLysValGlyArgProGlyArgLysArgLysArgLysArgLysArg 60  
 Db 365 ACGGCTCGGAGGTGGGAGGCTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 424  
 Qy 61 AspThrProLysAspProAlaValThrLysSerGlnProMetAlaGlnAspSerGly 80  
 Db 425 GACACCCCAAGGACCGAGGATGACCAAGTCTCAGCCCATGCGCCAGGACTCTGGC 484  
 Qy 81 ProSerAspLeuLeuProAsnGlyAspLeuGluLysArgSerGluProGluGlu 100

Db 485 CCTCGATCTCTACCCATGGAGACTTGGAGAGCGGAGTGAACCCCAACCTGAGGAG 544  
 Qy 101 GlySerProAlaAlaGlyGlnLysGlyGlyAlaProAlaGluGlyGlyGlyThrGluThr 120  
 Db 545 GGAGGCGGAGCTGCGAGGCGAAGAGGGTGGGGCCAGCTGAAGAGAGGGGAACCTGAGACC 604  
 Qy 121 ProProGluAlaSerArgAlaValGluAsnGlyCysCysValThrLysGluGlyArgGly 140  
 Db 605 CCACCAAGAGCCCTCAGAGCTGTGGGAATGGCTGTGTGTGACCAAGGAAGGCGGTGGA 664  
 Qy 141 AlaSerAlaGlyGluGlyLysGluGlnLysGlnThrAsnLysGlnLysSerMetLysMetGlu 160  
 Db 665 GCCTCTGCGAGGAGGCGCAAGAACAGAGCCCAACATCGAATCATGAAATGGAG 724  
 Qy 161 GlySerArgGlyArgLeuArgGlyGlyLeuGlyTyrGluSerSerLeuArgGlnArgPro 180  
 Db 725 GGCTCCCGGGCGGAGCTGCGAGGTGGCTTGGGCTGGAGTCCAGCTCCGTCAGGACCC 784  
 Qy 181 MetProArgLeuThrPheGlnAlaGlyAspProTyrTyrLysSerLysArgLysArgAsp 200  
 Db 785 ATGCCAAGACTCACCTTCCAGGCGAGGAGCCCTACTACTACGCAACAGGAACGGGAT 844  
 Qy 201 GluTyrLeuAlaArgTyrArgGluAlaGluLysLysAlaLysValLysAlaValMet 220  
 Db 845 GAGTGGCTGGCAGCTTGGAAAGGGAGGCTGAGAGAAAGCAAGGTAATTCAGTAATG 904  
 Qy 221 AsnAlaValGluGluAsnGlnAlaSerGlyGluSerGlnLysValGluGluAlaSerPro 240  
 Db 905 AATGCTGTGGAAGAAACCAAGGCTCTCGGAGAGTCTCAGAAAGGTGGAGGAGGCGCCCT 964  
 Qy 241 ProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrThrProGluPro 260  
 Db 965 CTTGCTGTGCAGCAGCCCGAGCCCTGCTTCTCCAGCTGTGGCCACCCCTCGAGCCA 1024  
 Qy 261 ValGlyLysAlaGlyAspLysAsnAlaThrLysAlaAlaAspAspGluProGluTyr 280  
 Db 1025 GTAGGAGGSGATGCTGGGGCAAGATGCTACCAAGCAGCCGAGTATGAGCTGAGTAT 1084  
 Qy 281 GluAspGlyArgGlyPheGlyLysGlyGluLeuValTyrGlyLysLeuArgGlyPheSer 300  
 Db 1085 GAGGATGGCGGGCTTTGGCATTTGGAGAGCTGGTGTGGGGGAAACTTTCGGGGCTTCTCC 1144  
 Qy 301 TrrTrrProGlyArgLysValSerTrrTrrMetThrGlyArgSerArgAlaAlaGluGly 320  
 Db 1145 TGGTGGCCAGGCGCAATTTGTCTTGTGTGATGACAGCCGAGCCGAGCAGCTGGAAGGC 1204  
 Qy 321 ThrAlaTrrValMetTrrPheGlyAspGlyLysPheSerValValCysValGluLysLeu 340  
 Db 1205 ACTCGCTGGGTCTGTGGTTCGGAGATGGCAAGTCTCAGTGGTGTGTGTGGAGAACTC 1264  
 Qy 341 MetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnProMet 360  
 Db 1265 ATGCCCTGAGCTCTCTCTGAGTGTGATTCACAGCCAGCCAGCTACCAACAGCAGCCCATG 1324  
 Qy 361 TyrArgLysAlaLysTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPhe 380  
 Db 1325 TACCGCAAGGCGCATCTACGAAGTCTCCAGGTGGCCAGCAGCCGTGCCCGGAAAGCTGTT 1384  
 Qy 381 ProAlaCysHisAspSerAspGluSerAspSerGlyLysAlaValGluValGlnAsnLys 400  
 Db 1385 CCAGCTTGCATGACAGTGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 444  
 Qy 401 GlnMetLysGluTrrAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLysGluPro 420  
 Db 1445 CAGATGATTTGAATGGGCGCTCGGTGGCTTCCAGCCCTCGGGTCTTAAGGGCTTGAACCA 1504  
 Qy 421 ProGluGluGluLysAsnProTyrLysGluValTyrThrAspMetTrrValGluProGlu 440  
 Db 1505 CCAG 1564  
 Qy 441 AlaAlaAlaTyrAlaProProProProAlaLysLysProArgLysSerThrThrGluLys 460

Db 1565 GCAGCTGCTTACGCCCAACCCCAACGCAAGAAACCCGAAAGAGACACAGAGAAA 1624  
 QY 461 ProlyValIysGluIleIleAspGluArgThrArgGluArgLeuValTyrGluValArg 480  
 Db 1625 CTTAAGGTCAAGAGATCATTTGATGAGCGCACAAAGGAGCGGCTGGTGTATGAGGTGCC 1684  
 QY 481 GlnLysCysArgSerIleGluAspIleCysIleSerCysGlySerLeuAsnValThrLeu 500  
 Db 1685 CAAAGGTGCAAGAAACATCGAGGACATTTGTATCTCATGTGGAGCCCTCAATGCACCCCTG 1744  
 QY 501 GlnHisProLeuPheIleGlyIleMetCysGlnAsnCysLysAsnCysPheLeuGluCys 520  
 Db 1745 GACACACCACTCTTCATTTGAGGAGCATGTGCAGAACTTAAAGAACTCTTCTTGAGTGT 1804  
 QY 521 AlaTyrGlnTyrAspAspArgTyrGlnSerTyrCysThrIleCysCysGlyArg 540  
 Db 1805 GCTTACAGATATACGACGATGGGTACCATCTCATTTGCACCATCTGTGGGGGGCGT 1864  
 QY 541 GluValLeuMetCysGlnAsnAsnAsnCysCysArgCysPheCysValGluCysValAsp 560  
 Db 1865 GAAGTGTCTATGTGTGGAAACAACCTGCGAGGTGCTTTGTGTCGAGTGTGTGAT 1924  
 QY 561 LeuLeuValGlyProGlyAlaIleGlnAlaIleIleGluAspProTyrPheAspTyr 580  
 Db 1925 CTCTTGTGGGCGCAGAGCTGCTCAGGACCATTTAAGAAAGCCCTGGAAGTGTAC 1984  
 QY 581 MetCysGlyHisIysGlyThrTyrGlyLeuLeuArgArgGluAspTyrProSerArg 600  
 Db 1985 ATCTGGCGGATAAAGGCACTATGGGCTCTCGAAAGAGGGAAGCTGGCTTCTCGA 2044  
 QY 601 LeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProlyValTyrPro 620  
 Db 2045 CTCAGATGTTCTTTGGCAATTAACATGACAGAAATTTACCCCAAGTTTACCCA 2104  
 QY 621 ProValProIleGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIleAla 640  
 Db 2105 CCTGTGCACTGAGAAAGAGAGGCCATCGGCTGTCTCTTGTGATGGGATGGT 2164  
 QY 641 ThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSerGlu 660  
 Db 2165 ACAGGGCTCTGTGCTGGAAGGAGCTGGGATCCAGATGAGACCGCTACATGCTCCGAG 2224  
 QY 661 ValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyIleMetTyrVal 680  
 Db 2225 GTGTGTGAGACTTCATCAGGTGGGATGTGTGGGACACAGGAAAGATCATGTACGTC 2284  
 QY 681 GlyAspValArgSerValThrGlnLysHisIleGlnGluTyrProIleAspLeuVal 700  
 Db 2285 GGGGACCTCGCAGCGTCACACAGAACATATCCAGAGTGGGGCCCATTCGACCTGGTG 2344  
 QY 701 IleGlyGlySerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeuTyr 720  
 Db 2345 ATTTGAGGAGCTCCCTGCATGTACCTTCATTTGTCAACCTGCCCCGCAAGGACCTTAT 2404  
 QY 721 GluGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgProLys 740  
 Db 2405 GAGGGTACTGGCGGCTCTTCTTGAAGTTCACGCGCTCTCGATGATGCGCGGCCAAG 2464  
 QY 741 GluGlyAspAspArgProPhePheThrPheLeuPheGluAsnValAlaIleMetGlyValSer 760  
 Db 2465 GAGGGAATATGCGCCCTTCTTGTGCTCTTGAAGATGTGGGGCCATGGGGGTAGT 2524  
 QY 761 AspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLysGlu 780  
 Db 2525 GACAAAGGAGACATCTGCGATTTCTTGAGTCTAACCCCGTATGATGACCCAAAGAA 2584  
 QY 781 ValSerAlaIleHisArgAlaArgTyrPheTyrPheLysLeuLeuProGlyMetAsnArgPro 800  
 Db 2585 GTGTCTCTGCACACAGGCGCGGTACTCTGGGGTAACCTCTCGCATGAACAGGCT 2644  
 QY 801 LeuAlaSerThrValAsnAspLysLeuGluGlnGluCysLeuGluHisGlyValTyrIle 820  
 Db 2645 TTGGCATCCACTGTGAATGATTAAGCTGGAAGTGTGAGAGCGGACAGGAGATA 2704

QY 821 AlaLysPheSerLysValArgThrIleThrArgSerAsnSerIleLysGlnGlyLys 840  
 Db 2705 GCCAAGTTCAGCAAGAGTGAAGACATTACACAGGTCAAACTCTTAAGACGGGCAAA 2764  
 QY 841 AsnGlnHisPheProValPheMetLeuGluLysGluAspIleLeuTyrCysThrGluMet 860  
 Db 2765 GACACACATTTCCCGTCTTCATGAACGAGAGAGACATCTGTGTGCATGTAATG 2824  
 QY 861 GluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeuAla 880  
 Db 2825 GAAAGGTGTGTGGCTTCCCGTCACCTACAGAGCTTCCAACTGAGCCGCTGGGG 2884  
 QY 881 ArgGlnArgLeuLeuGlyArgSerTyrSerValProValIleArgHisLeuPheAlaPro 900  
 Db 2885 AGGACAGAGACTGTGGCGCATGTGAGACGTGTGCGGTATCGGCCACTCTTGCTCCG 2944  
 QY 901 LeuLysGluTyrPheAlaCysVal 908  
 Db 2945 CTGAAGAAATATTTGCTGTGTG 2968  
 RESULT 3  
 AF067972 3005 bp mRNA linear PRI 12-FEB-2001  
 LOCUS Homo sapiens DNA cytosine methyltransferase 3 alpha (DNMT3A) mRNA,  
 DEFINITION complete cds.  
 ACCESSION AF067972  
 VERSION AF067972.2 GI:12746531  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE  
 AUTHORS Xie,S., Wang,Z., Okano,M., Nogami,M., Li,Y., He,W.W., Okumura,K.  
 and Li,E.  
 TITLE Cloning, expression and chromosome locations of the human DNMT3  
 gene family  
 JOURNAL Gene 236 (1), 87-95 (1999)  
 MEDLINE 99365304  
 PUBMED 10433969  
 REFERENCE  
 2 (bases 1 to 3005)  
 AUTHORS Xie,S., Okano,M. and Li,E.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-MAY-1998) CVRC, Mass. General Hospital, 149 13th  
 Street, Charlestown, MA 02129, USA  
 3 (bases 1 to 3005)  
 REFERENCE  
 Hata,K., Shirohzu,H., Sasaki,H. and En,L.  
 AUTHORS Direct Submission  
 TITLE Submitted (12-FEB-2001) CVRC, Mass. General Hospital, 149 13th  
 Street, Charlestown, MA 02129, USA  
 REMARK  
 COMMENT Sequence update by submitter  
 On Feb 12, 2001 this sequence version replaced gi:4927369.  
 FEATURES  
 source  
 1..3005  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2P23: between D2S171 and D2S174; 48-50 cM"  
 1..3005  
 /gene="DNMT3A"  
 230..2968  
 /gene="DNMT3A"  
 /function="de novo DNA cytosine methyltransferase"  
 /codon\_start=1  
 /product="DNA cytosine methyltransferase 3 alpha"  
 /protein\_id="AAD33084.2"  
 /db\_xref="GI:12746532"  
 /translation="MPAMPSSGGPPTSSAAREEDRKDGEDEPRGKEERQBPST  
 AKVGRDPRKRPVPSGDPKDAVUSKPSMODSGASRLLPNGLEKRSRPOPE  
 EESPAGGQKGAFAEGEALETLPASAAVENGGCTPRGGRAPPAEGEKEQETNTES  
 NMWESRRRLRGGLGWESSLRQPMPLRTPQAGDGYISKRRDMLARWKEAEKKA  
 KVIAQMAVNEVNGGSEQKVEASPPAVOQPTDASPTVATTPPEVGSADADKNATK

AGDDEPEYEDGRGFGIGELVGMKLRGFSWMPGRIVSIWMNTRSRRAAEQTRVWVWFQDG  
KFSVVCVEKLMPLSFSFSAHQATYNKOPMYRKATIEYVLOVASSRAGKLPVCHDSDE  
SDTAKAVEONKPMIEMALGGFQSPGKLEPEPEEKNPKEVYTDMMVPEAAAVAP  
PPAKKPRKSTAEKPKVKEIIDERTRERLYEVROKCRNIEDI CISCGLNVTLEHPL  
FVGMQCNKCNFLCECAVQDDDGVSQYCTICCGREVLMCGNNNCRCCFCVCDLL  
VGPQAQAALKEDEPNVCYMGCHGTGVLRRREDMPSLQMFANNDHDSFDPKVPY  
PVAEKRPKIRVLSIFDGLIATGLLVLDGLOVDVYIASEVEDSITVMVRHOGKIM  
YVGDVSRTOKHIOEMGFDFLVIGGSPCDLSIVNPASKGLYEGTGRLFFRFVRLIHD  
ARPKGDDRPFFWLFEVNAVANGSDREDISRFLESNPMDAKVUSAHAHARFPWNL  
PGNRRFEMERVFVGFPHVTIDVNSNMRKLRQLLGRSWSVPVIRHLFAPLKEYFACV-  
BASE COUNT 674 a 846 c 978 g 507 t

ORIGIN

## Alignment Scores:

Pred. No.: 2,48e-203 Length: 3005  
Score: 4729.50 Matches: 875  
Percent Similarity: 97.14% Conservativity: 8  
Best Local Similarity: 96.26% Mismatches: 25  
Query Match: 96.34% Indels: 1  
DB: 9 Gaps: 1

US-09-720-086-5 (1-908) x AF067972 (1-3005)

Qy 1 MetProSerGlyProGlyAspThrSerSerLeuGluArgGluAspArg 20  
Db 239 ATGCCCTCCAGCGGCCCGGGACACACAGCAGCTCTGCTCGGAGCGGGAGGACCGA 298  
Qy 21 LysGluGlyGluGluGluGluAsnArgGlyLysGluGluGluGluGluProSerAla 40  
Db 299 AAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 358  
Qy 41 ThrAlaArgLysValGlyArgProGlyArgLysArgLysHisProValGluSerSer 60  
Db 359 ACGCAGCAGAAAGTGGGGCGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 418  
Qy 61 AspThrProLysAspProAlaValThrThrLysSerGlnProMetAlaGlnAspSerGly 80  
Db 419 GACACGCCAAGACCCCTCGGTGTATCTCCAAAGTCCCATCCATGTCGCCAGGACTCAGGC 478  
Qy 81 ProSerAspLeuLeuProAsnGlyAspLeuLysArgSerGluProGlnProGluGlu 100  
Db 479 GCCTCAGAGCTATTATCCCAATGGGACTTGGAGAGGCGGAGTGGAGCCCGCAGGAGGAG 538  
Qy 101 GlySerProAlaAlaGlyGlnLysGlyGlyAlaProAlaGluGlyGluGly---ThrGlu 119  
Db 539 GGGAGCCCTGCTGGGGGCGAAGGCGGGCGGCCCGCCAGCAGAGGAGGAGGAGTGCAGCTGAG 598  
Qy 120 ThrProGluAlaSerArgAlaValGluAsnGlyCysValThrLysGluGlyArg 139  
Db 599 ACCCTGCTGAGCCCTCAGAGCAGTGGAAATGGCTGCTGCCACCCCAAGGAGGAGGAGGAG 658  
Qy 140 GlyAlaSerAlaGlyGluGlyLysGluGlnLysGlnThrAsnLeuGluSerMetLysMet 159  
Db 639 GGAGCCCTCGAAGCGGGCGCAAGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 718  
Qy 160 GluLysSerArgGlyArgLeuArgGlyLysLeuGlyLysLeuGlyLysLeuGlyLys 179  
Db 719 GAGGGCTCCCGGGCGGGCTGGCGGGTGGCTTGGGGTGGAGTCCAGGCTCCGTCAGGCGG 778  
Qy 180 ProMetProArgLeuThrPheGlnAlaGlyAspProTyrTyrLysSerLysArgLysArg 199  
Db 779 CCCATGCGAGGCTACCTTCAGGCGGGGAGACCCCTACTACATCAGCAAGCGCAGGCGG 838  
Qy 200 AspGluTrpLeuAlaArgTrpLysArgGluAlaGluLysLysAlaLysValIleAlaVal 219  
Db 839 GACGAGTGGTGGCGCCCTGGAAAAGGGAGGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 898  
Qy 220 MetAsnAlaValGluGluAsnGlnAlaSerGlyGluSerGlnLysValGluGluAlaSer 239  
Db 899 ATGAATGCTGTGGAAGAAACACAGGGGCGGGGAGTCTCAGAAAGTGGAGGAGGAGGAGC 958  
Qy 240 ProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrThrProGlu 259

Db 959 CCTCTGCTGTGACAGCAGCCCACTGAGCCCGCATCCCCCACTGTGTGCTTACCAACGCTGAG 1018  
Qy 260 ProValGlyGlyAspAlaGlyAspLysAsnAlaThrLysAlaAlaAspAspGluProGlu 279  
Db 1019 CCGTGGGGTCCGATGCTGGGGGCAAGATGCCCAAGCAGGCGATGACAGCAGCCAGAG 1078  
Qy 280 TyrGluAspGlyArgGlyPheGlyIleGlyGluLeuValTrpGlyLysLeuArgGlyPhe 299  
Db 1079 TAGCAGACGGCGGGGCTTTGGCATTTGGGAGCTGTGTGGGGGAAACTCGGGGGCTTC 1138  
Qy 300 SerTrpTrpProGlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGlu 319  
Db 1139 TCCTGTGTGCCAGCGCCCATTTGTCTTGTGTGATACGGGCGGAGCCGAGCAGCTGAA 1198  
Qy 320 GlyThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLys 339  
Db 1199 GGCACCCGCTGGGTCTATGTGTTTCGAGAGCGGCAATTTCTCAGTGGTGTGTGTGAGAAG 1258  
Qy 340 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnPro 359  
Db 1259 CTGATCCGCTCAGCTCGTTTTCAGTGTCTCCACAGGCCACGTACAACAAGCAGCC 1318  
Qy 360 MetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu 379  
Db 1319 ATGTACCGCAAGACCATCTACAGGTCTCTGAGGTGGCCAGCAGCGCGGGGAGCTG 1378  
Qy 380 PheProAlaCysHisAspSerAspGluSerAspSerGlyLysAlaValGluValGlnAsn 399  
Db 1379 TTCCCGGTGTCCACAGCAGCGATGAGGTGACACTGCCAAGCGGTGGAGGTGCAGAAC 1438  
Qy 400 LysGluMetIleGluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGlu 419  
Db 1439 AAGCCCATGATTGAATGGCCCTGGGGGGCTTCCAGCCTTCTGGCCCTTAAGGGGCTGGAG 1498  
Qy 420 ProProGluGluGluLysAsnProTyrLysGluValTyrThrAspMetTrpValGluPro 439  
Db 1499 CCACCAAGAAAGAGAAATCCCTACAAAGAGTGTACACGACATGTGGGTGGAACCT 1558  
Qy 440 GluAlaAlaTyrAlaProProProAlaLysLysProArgLysSerThrThrGlu 459  
Db 1559 GAGCAGCTGCTCAGCAGCAGCTCCACAGCCCAAAAGCCCGGAGAGCAGCAGCGAG 1618  
Qy 460 LysProLysValLysGluIleAspGluArgThrArgGluArgGluValTyrGluVal 479  
Db 1619 AAGCCCAAGGTCAAGGAGATTATTGATGAGCGCACAGAGAGCGGCTGGTGTACGAGGTG 1678  
Qy 480 ArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThr 499  
Db 1679 CGCAGAGTGCAGGACATTTGAGGACATCTGCATCTCTGTGGAGGCTCAATGTTACC 1738  
Qy 500 LeuGluHisProLeuPheIleGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGlu 519  
Db 1739 CTGGAACACCCCTCTTCTGTGGAGGAATGTGCCAAACTGCAAGAACTGCTTTCTGGAG 1798  
Qy 520 CysAlaTyrGlnTyrAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyGly 539  
Db 1799 TGTGCGTACAGTACGAGCAGCGGCTACCACTCTACTCACCATCTGCTGTGGGGC 1858  
Qy 540 ArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysVal 559  
Db 1859 CGTGAGGTCTCATGTGCGGAAACAACACTGCTGTCAGGTGCTTTTGGTGGAGTGTGTG 1918  
Qy 560 AspLeuLeuValGlyProGlyAlaAlaGlnAlaIleLysGluAspProTrpAsnCys 579  
Db 1919 GACCTCTTGGTGGGCGGGGGCTGCCAGGAGGACCATTAAGGAAGACCCCTGGAACCTGC 1978  
Qy 580 TyrMetCysGlyHisLysGlyThrTyrGlyLeuLeuArgArgGluAspTrpProSer 599  
Db 1979 TACATGTGCGGCAAGGGTACTTACGGGTGCTCGGGGCTGCTCGCGCGGAGAGGAGTGGCCCTCC 2038  
Qy 600 ArgLeuGlnMetPhePheAlaAsnHisAspGlnGluPheAspProProLysValTyr 619

```

Db 2039 CGGCTCCAGATGTTCTTCGCTTAATTAACACAGCAGCAAGATTTCACCTCCAAAGCTTTAC 2098
Qy 620 ProProValProAlaGluValArgLysProIleArgValLeuSerLeuPheAspGlyIle 639
Db 2099 CCACTGCTCCACACTGAGAAAGAGAGCCATCCGGGTGCTGCTCTTGTGATGGAATC 2158
Qy 640 AlrThGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSer 659
Db 2159 GCTACAGAGGCTCTGCTGCTGAGAGACTTGGGCATTCAGCTGAGCCGCTCATTTGCTCG 2218
Qy 660 GluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyr 679
Db 2219 GAGGTGTGTAGAGACTTCATCAGCTGGGCGATGTGGGCAACAGGGGAAGATCATGTAC 2278
Qy 680 ValGlyAspValArgSerValThrGlnLysHisIleGlnGlnLutProGlyProPheAspLeu 699
Db 2279 GTGCGGGAGCTCGGACGCTCACAGAGACATATCCAGAGTGGGGCCATTCGATCTG 2338
Qy 700 ValIleGlyGlySerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu 719
Db 2339 GTGATTGGGGGAGCTCCCTGCAATGACTCTCCATCGTCAACCTGCTCGCAAGGCGCTC 2398
Qy 720 TyrGlnGlyThrGlyArgLeuPhePheGluPheThrArgLeuLeuHisAspAlaArgPro 739
Db 2399 TACGAGGACACTGACCGGCTCTTGTAGATCTACCGCTCTCGCATGATGCGGCGCC 2458
Qy 740 LysGlnGlyAspAspArgProPhePheTrpLeuPheGlnAsnValValAlaMetGlyVal 759
Db 2459 AAGGAGGAGATGATCGCCCTCTTCTTGCTGCTCTTGAGAAATGGTGGGCGCATCGCGCTT 2518
Qy 760 SerAspLysArgAspIleSerArgPheLeuGlnSerAsnProValMetIleAspAlaLys 779
Db 2519 AGTGACAAAGGAGACATCTCGCATTTCTCGAGCCCAACCTGTGATGATGACCAAA 2578
Qy 780 GluValSerIleAlaHisArgAlaArgTyrPheThrGlyLysAsnLeuProGlyMetAsnArg 799
Db 2579 GAGGTGTGACTGACACAGAGGCGCGCTACTTCTGGGGTAACTCTCCGGTATGAAACAGG 2638
Qy 800 ProLeuAlaSerThrValAsnAspLysLeuGlnLysCysLeuGlnLysIleGlyValArg 819
Db 2639 CCGTTGGCATCCACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2698
Qy 820 IleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGly 839
Db 2699 ATAGCCAAAGTTACAGAAAGAGAGACCATTAACAAGAGTCAAACTCCATTAAGACAGGCG 2758
Qy 840 LysAspGlnHisPheProValPheMetAsnGlnLysGlnAspIleLeuTyrCysThrGln 859
Db 2759 AAGAGCCAGCATTTTCTGCTCTCATGATGAGAAAGAGACATCTTATGATGATGATGAA 2818
Qy 860 MetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeu 879
Db 2819 ATGGAAGAGGATATTTGGTTTCCAGTCCACTATATCTGACTTCCAAACATGAGCGGCTTG 2878
Qy 880 AlaArgGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAla 899
Db 2879 GCGAGGAGAGAGACTGCTGGGCGGCTGATGAGCGTGCACAGTCAATCCGCCACTTCTGCT 2938
Qy 900 ProLeuLysGlnLysPheAlaCysVal 908
Db 2939 CCGCTGAAGAGATATTTTGGCTGTGTG 2965

RESULT 4
AF331856 4258 bp mRNA linear PRI 02-JAN-2002
LOCUS AF331856 Homo sapiens DNA cytosine methyltransferase 3 alpha (DNMT3A) mRNA,
DEFINITION complete cds.
ACCESSION AF331856
VERSION AF331856.1 GI:18033252
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
AUTHORS 1 (bases 1 to 4258)
TITLE Cloning, expression and characterization of human DNMT3 genes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4258)
AUTHORS Ni, J., Pradhan, S. and Roberts, R.J.
TITLE Direct Submission
JOURNAL Submitted (22-DEC-2000) New England Biolabs, 32 Tozer Road,
Beverly, MA 01915, USA

FEATURES
source location/Qualifiers
1..4258
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
1..4258
/gene="DNMT3A"
227..2956
/gene="DNMT3A"
/codon_start=1
/product="DNA cytosine methyltransferase 3 alpha"
/protein_id="AA157039.1"
/db_xref="GI:18033253"
/translation="WPSGPGDTSSGALEEREDRKDGEERPRGKERQERESTTARK
VGRPRKRKHPPVESGDTPRKDAVISKSPMAQDSAGSELPLNGLEKRSERQPEGS
PAGGQGGAPAGEGAETLPEASRAVANGCTPRKEGAPAEAKKEKNTIESRM
EGSRRLGLGAGMESLRLRPMPLRTPFAGPBYISKKRDEMLRMKREAKKVI
AGMNVERNOGFESOKVEFASPPAVOFTPTSPPTVATTPPEVSDGDKNATKAG
DEPEREDRGRIGELVNGKLRGFSWMPGRIVSWMTGSRSAEETRRYMPGDKFS
VVCERKLMPLSFSCAFHQTATNKQPMTRKALYEVLQVASSRAGLFFVCHDSDSDT
AKAVEVQKPMIEMALGGFOPSGRGLSEPRBEENKPYELVTDMMVEPAAVAPP
AKPRKSTAEKPKVKEIIDERTRELVEVOKCNIEDICISGSLVLEHPLFVG
GMCONKRCFLFCAYOYDDGYOSTICCGREVLNMGNNNCRFCVECVDLVPG
GAOAKAIKEDPMNCGHKGKGYGLRREBPSRLQMFANHDOERDPVYPPV
AEKRPITVLSTFQSLATGLVTKDGIQVDRYASRVCEDSITYGMRHOCKIYVG
DVRSTQKHIOBMPFDLVITGSPENDSINYPAKGLVEGTGRIFPEPYRLHARP
KEGDRPFPFLVEENVAVAGVSDKRDISFLSNPWIADKEVSAHARARYEWGILPM
NRPLASTVNDKLELOECLEHGRIAKFSKVRITTSNSIKQKQDFVFNMEKEDIL
WTEBERVFGFVHYTVDVSNMSRLRORLGRSWSVPIRHIFALKEFYACV"

BASE COUNT 1062 a 1089 c 1257 g 850 t
ORIGIN

Alignment Scores:
Pred. No.: 3..54e-203 Length: 4258
Score: 4729.50 Matches: 875
Percent Similarity: 97.14% Conservative: 8
Best Local Similarity: 96.26% Mismatches: 25
Query Match: 96.34% Indels: 1
DB: Gaps: 1

US-09-720-086-5 (1-908) x AF331856 (1-4258)
Qy 1 MetProSerSerGlyProGlyAspThrSerSerSerLeuGlnArgIleAspAspArg 20
Db 227 ATGCCCTCCAGCGCGCCCGGAGACACAGCTGCTGCGGAGCGGAGGAGGCCA 286
Qy 21 LysGlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 40
Db 287 AAGGACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 346
Qy 41 ThrAlaArgLysValGlyArgProGlyArgLysArgLysHisIleProProValGluSerSer 60
Db 347 ACGGACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 406
Qy 61 AspThrProLysAspProAlaValThrThrLysSerGlnProMetLysGlnAspSerGly 80
Db 407 GACACGCCAAGAGACCTCGGAGATCTCCAGTCCCATTCATGCGCCAGGACTCAGGC 466
Qy 81 ProSerAspLeuLeuProAsnGlyAspLeuGlnLysArgSerGlnProGlnProGlnGln 100
Db 467 GCCTCAGAGCTATTACCAATGGGACTTGAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGG 526
Qy 101 GlySerProAlaAlaGlnGlnLysGlyValAlaProAlaGlnGlnGlnGlnGlnGlnGln 119

```





Qy 840 LysAspGlnHisPheProValPheMetAsnGlnLysGlnAspIleLeuTrpCysThrGlu 859  
 Db 2747 AAGACGACGATTTCTCTCTCATGAAAGAGAAAGACATCTATGTCACACGAA 2806  
 Qy 860 MetGlnArgValPheGlyPheProValHisTyThrAspValSerAsnMetSerArgLeu 879  
 Db 2807 ATGGAAAGGATATTGTTCCAGTCACCTAATACCTCTCCCAACATGAGCCCGCTTG 2866  
 Qy 880 AAlaArgGlnArgLeuGluArgSerTrpSerValProValIleArgHisLeuPheAla 899  
 Db 2867 GCGAGCGACAGACTGCTGGCCGCTCATGAGCGCTGCATCTATCCGCCACTCTTCGCT 2926  
 Qy 900 ProLeuLysGlnTyrrPheAlaCysVal 908  
 Db 2927 CCGCTGAGAGATATTGCGTGTGTG 2953  
 REF 5  
 AC 2924  
 LOCUS  
 DEFINITION Rattus norvegicus clone CH230-282D16, \*\*\* SEQUENCING IN PROGRESS  
 AC120824 123936 bp DNA linear HTG 23-JUL-2002  
 AC120824  
 AC120824.2 GI:21902541  
 HTG: HTG5\_PHAS1.  
 KEYWORDS Rattus norvegicus.  
 SOURCE Rattus norvegicus.  
 ORGANISM Rattus norvegicus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 123936)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
 Barbieri,J., Benton,J., Bimaga,K., Blankenburg,K., Bonini,T.,  
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Cartton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Barnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
 Gabali,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gorielli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
 Hombl,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louieged,H.,  
 Lozad,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Masehwar,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
 Massey,E., Mawney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogun,R., Okunom,G.,  
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
 Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,  
 Scherer,S., Scott,G., Shen,H., Shooshtari,N., Stenson,I.,  
 Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H.,  
 Suton,A., Svatek,A., Tabot,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Tansay,U., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,  
 Williams,G., Williams,A., Wlezyk,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstein,G., and Gibbs,R.  
 TITLE  
 JOURNAL  
 Direct Submission  
 Unpublished

REFERENCE  
 2 (bases 1 to 123936)  
 AUTHORS  
 TITLE  
 JOURNAL  
 DIRECT SUBMISSION  
 SUBMITTED (13-MAY-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 123936)  
 AUTHORS  
 TITLE  
 JOURNAL  
 DIRECT SUBMISSION  
 SUBMITTED (23-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 ON JUL 18, 2002 this sequence version replaced gi:20531783.  
 COMMENT  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 Project Information  
 Center project name: GMLV  
 Center clone name: CH230-282D16  
 Summary Statistics  
 Sequencing vector: Plasmid:  
 Chemistry: Dye-terminator Big Dye 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 95907 bases at least Q40  
 Consensus quality: 99223 bases at least Q30  
 Consensus quality: 101036 bases at least Q20  
 \*\*\*\*\*  
 NOTE: Estimated insert size may differ from sequence length  
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 38 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.  
 1  
 1009: contig of 1009 bp in length  
 1010: gap of unknown length  
 1110: contig of 1004 bp in length  
 2113: gap of unknown length  
 2213: gap of unknown length  
 2214: contig of 1020 bp in length  
 3233: gap of unknown length  
 3234: contig of 1109 bp in length  
 4442: gap of unknown length  
 4443: contig of 1660 bp in length  
 4543: gap of unknown length  
 6202: contig of 1162 bp in length  
 6203: gap of unknown length  
 7464: gap of unknown length  
 7465: contig of 1208 bp in length  
 7565: gap of unknown length  
 8773: contig of 1499 bp in length  
 8773: gap of unknown length  
 10371: contig of 1243 bp in length  
 10372: gap of unknown length  
 10471: contig of 1243 bp in length  
 11714: gap of unknown length  
 11715: contig of 1110 bp in length  
 11815: gap of unknown length  
 12924: contig of 1110 bp in length  
 12924: gap of unknown length  
 13024: gap of unknown length  
 13025: contig of 1333 bp in length  
 14347: gap of unknown length  
 14348: gap of unknown length  
 15652: contig of 1205 bp in length  
 15653: gap of unknown length  
 15752: contig of 2044 bp in length  
 15753: gap of unknown length  
 17796: contig of 1041 bp in length  
 17797: gap of unknown length  
 17896: gap of unknown length  
 18937: contig of 1041 bp in length  
 18938: gap of unknown length  
 19037: gap of unknown length  
 21543: gap of 2506 bp in length  
 21544: gap of unknown length  
 21544: contig of 2897 bp in length  
 24540: gap of unknown length  
 24541: gap of unknown length  
 24641: contig of 3472 bp in length  
 28112: gap of unknown length  
 28113: gap of unknown length

```

* 28213 31101: contig of 2889 bp in length
* 31102 31201: gap of unknown length
* 31202 34415: contig of 3214 bp in length
* 34416 34515: gap of unknown length
* 34516 37353: contig of 2838 bp in length
* 37354 37453: gap of unknown length
* 37454 40938: contig of 3485 bp in length
* 40939 41038: gap of unknown length
* 41039 44331: contig of 3293 bp in length
* 44332 48035: contig of 3604 bp in length
* 48036 48135: gap of unknown length
* 48136 51181: contig of 3046 bp in length
* 51182 51281: gap of unknown length
* 51282 55361: contig of 4080 bp in length
* 55362 55461: gap of unknown length
* 55462 59580: contig of 4119 bp in length
* 59581 59680: gap of unknown length
* 59681 63525: contig of 3845 bp in length
* 63526 67434: contig of 3809 bp in length
* 67435 67534: gap of unknown length
* 67535 72194: contig of 4660 bp in length
* 72195 72294: gap of unknown length
* 72295 77777: contig of 5383 bp in length
* 77778 81349: contig of 3572 bp in length
* 81350 86396: contig of 4947 bp in length
* 86397 91224: gap of unknown length
* 91225 91324: contig of 4728 bp in length
* 91325 97693: contig of 8445 bp in length
* 97694 97869: gap of unknown length
* 97870 105670: contig of 7801 bp in length
* 105671 105770: gap of unknown length
* 105771 111368: contig of 5598 bp in length
* 111369 111468: gap of unknown length
* 111469 116680: contig of 5212 bp in length
* 116681 116780: gap of unknown length
* 116781 123936: contig of 7156 bp in length.

```

## FEATURES

```

source
1. .123936
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-282D16"

```

```

BASE COUNT 35904 a 21413 c 21468 g 36486 t 8665 others
ORIGIN

```

## Alignment Scores:

```

Pred. No.: 3 62e-105 Length: 123936
Score: 2571.50 Matches: 487
Percent Similarity: 92.23% Conservative: 23
Best Local Similarity: 88.07% Mismatches: 27
Query Match: 52.38% Indels: 18
DB: 2 Gaps: 4

```

```

US-09-720-086-5 (1-908) x AC120824 (1-123936)

```

```

Qy 203 LeuAlaArgTrpLysArgGluAlaGluLysAlaLysValLysValLysAlaValMetAsnAla 222
Db 84783 CTGGCCCTTATGG-----CTTGAGAGAAAGCCAGGTGATTGCAATAATCAATGCT 84833
Qy 223 ValGluGluAsnGlnAlaSerGlyGluSerGlnLysValGluGluAlaSerProAla 242
Db 84834 GTGAGAGGAAACAGGAGGCTCTGGGAGGCTCAGAGAGGTGGAGAGCCAGCCCTCTGCA 84893
Qy 243 ValGlnGlnProThrAspProAlaSerProThrValAlaThrProGluProValGly 262
Db 84894 GTGCAGAGAGCCAG-GACCCTGCATCTTCCACTGTGGCCATTACCCCTGTGCCAGTAGG 84952
Qy 263 GlyAspAlaGlyAspLysAsnAlaThrLysAlaAlaAspAspGluProGluLysThr 282

```

```

Db 84953 GCTGATGCTGGGCAAGAAATGCTATCAAGGCAGCTGATGAGGACCCCGAGTATGAGGAT 85012
Qy 283 GlyArgGlyPheGlyLysGlyLeuValTrpGlyLysLeuArgGlyPheSerTrpTrp 302
Db 85013 GGCAGGAGGCTTGGCATTTGGAGAGCTGTGTGGGGAATACTCCGGGGCTTCTCTGGTGG 85072
Qy 303 ProGlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGluGlyThrArg 322
Db 85073 CCAAGCCGAATTTGTCTTGTGTGATGACAGGCCAGCAGCAGCAGCGAAGGACCTCGC 85132
Qy 323 TrpValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLysLeuMetPro 342
Db 85133 TGGGTATGTTGGTGGAGATGGCAAGTTTTCAGTATGATGTGTGAGAAGCTCATGCC 85192
Qy 343 LeuSerSerPheCysSerAlaPheHisGlnAlaThrTrpAsnLysGlnProMetTrpArg 362
Db 85193 CTGAGCTCTTCTGCAT----TTCACACAGGCCACCTACCAACAGCAGCCATGTACCGC 85248
Qy 363 LysAlaIleTrpGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPheProAla 382
Db 85249 AAAGCCATCTACGAAGTCTCTCAGGTGGCCAGCAGCTGTCAGGGAGCTGTTCCACGCG 85308
Qy 383 CysHisAspSerAsp--GluSerAspSerGlyLysAlaValGluValGlnAsnLysGlnM 402
Db 85309 TGTATGACAGCAGCATGAAAGTGACACTGGCAAGCTGTGGAGGTGCAGAACCAAGCAGA 85368
Qy 402 etileGluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGluProProG 422
Db 85369 CGATTGAGTGGGTCTTGGCGGCTTCAGCCCTCTGGTCCCAAGGGCTGGAGCCACAG 85428
Qy 422 LuGluLysAsnProTrpLysGluValTrpThrAspMetTrpValGluProGluAla 442
Db 85429 AAGAGGAGAACAACTTTTACAGGAAGTTTACACCGACATGTGGGTGGAGCCTGAGGCAG 85488
Qy 442 laaIaTrpAlaProProProAlaLysLysProArgLysSerThrThrGluLys--Pro 461
Db 85489 CTCCTTATTTCCCAACCCCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 85539
Qy 462 LysValLysGluIleLysAspGluArgThrArgGluArgLysValTrpGluValArgGln 481
Db 85540 AAGGTCAAGGAGATCATTTGATGAACACACAAAGAAAGCAAGCAAGCAAGCAAGCAAG 85599
Qy 482 LysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThrLeuGlu 501
Db 85600 AAGTGCAAAACATCGAGGACATTTGTATCTCATGTGGAGCCCTCAATGTCCACCTGGAG 85659
Qy 502 HisProLeuPheIleGlyGlyMetCysGlnAsnCysAlaValAsnCysPheLeuGluCyeAla 521
Db 85660 CACCCACTCTTGTGTGGAAATGTGCCAGAACTGTNAGAACTGCTTCTTGGAGTGTCT 85719
Qy 522 TyrGlnTrpAspAspAspGlyTrpGlnSerTrpCysThrIleCysGlyGlyArgGlu 541
Db 85720 TACCATACGATGACGATGTGTACCAAGTCTTCTGTACATCTCTGTGGGGGCGGAA 85779
Qy 542 ValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysValAspLeu 561
Db 85780 GTGCTCATGTGTGGGAACAACAACCCGCTGAGGTGCTTTTGTGGAGTGTGTGGATCTC 85839
Qy 562 LeuValGlyProGlyAlaAlaGlnAlaIleLysGluAspProTrpAsnCysTrpMet 581
Db 85840 TTGGTGGGCGCAGGGGCTGGCCCAAGCAGCAGCATTAAGAGAGACCCCTGGAACTGTCTCATG 85899
Qy 582 CysGlyHisLysGlyThrTrpGlyLeuLeuArgArgArgGluAspTrpProSerArgLeu 601
Db 85900 TGT-----GAGCTGCTGCGGAGAGGAGGAGCTGGCCTTCTCGACTC 85941
Qy 602 GlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTrpProPro 621
Db 85942 CAGATGTTCTTTGGCAATAACACACAGCAGGAAATTTGACCCCGCAAGGTTTACCCACCT 86001
Qy 622 ValProAlaGluLysArgLysProIleArgValLeuLeuPheAspGlyIleAlaThr 641
Db 86002 GTGCCAGCTGAGAGAGAGAGCCCATCCAGGTGCTATCTCTATTATTTGATGGATTGTCTACA 86061

```

QY 642 GlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSerGluVal 661  
DB 86062 GGGCTCTGCTGTCGAAGGAGCCTGGGATCCAGTGCACCGCTACATTGCTCCGAGGTG 86121  
QY 662 CysGluAspSerIleThrValIleGlyMetValArgHisGlnIleLysIleMetCysValGly 661  
DB 86122 TGGAGGACTCCATCCACCGGGGCTATGTGGCACCAGGAGAAAGATCATGTGTGGGG 86181  
QY 682 AspValArgSerValThrGlnLysHisIleGlnIleTyrPheProPheAspLeuValIle 701  
DB 86182 GAGCTCGCAGCATCACACGAGAGCATATCCAGAGAGGGGCCATTCATCTGGTGATT 86241  
QY 702 GlyGlySerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeuTyrGlu 721  
DB 86242 GGGGGCAGTCTTGCATATGACCTCTCCATCATCAACCTCCGCAAGGAGCTTTACAG 86301  
QY 722 GlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgProLysGlu 741  
DB 86302 GGCACCTGGCAGCTCTTCTTGTGAGTTCTTACACCTCTCTGATATGCGGCCCAAGAG 86361  
QY 742 GlyAspAspArgProPhePheTyrLeuPheGlu 752  
DB 86362 GGAGAGATGCCCCCTTCTTGTGCTTTGAG 86394  
RESULT 6 AK025230 2191 bp mRNA linear PRI 29-SEP-2000  
LOCUS Homo sapiens cDNA: FLJ21577 fls, clone COL06724, highly similar to  
DEFINITION AF067972 Homo sapiens DNA cytosine methyltransferase 3 alpha  
(DNMT3A) mRNA.  
ACCESSION AK025230  
VERSION AK025230.1 GI:10437699  
KEYWORDS oligo capping; fls (full insert sequence).  
SOURCE Homo sapiens colon cDNA to mRNA, clone\_lib:COL clone:COL06724.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y.,  
Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T.,  
Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2191)  
JOURNAL Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T.,  
TITLE Shibahara, T., Tanaka, T. and Nakamura, Y.  
JOURNAL Direct Submission  
COMMENT Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure Analysis, Human  
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,  
Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
Fax:81-3-5449-5416)  
NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing: Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing: Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).  
FEATURES  
source Location/Qualifiers  
1. .2191  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="COL06724"  
/cissue\_type="colon"  
/clone\_lib="COL"  
/note="Cloning vector pME18SFL3"  
misc\_feature 1. .2191  
/note="highly similar to AF067972 Homo sapiens DNA  
cytosine methyltransferase 3 alpha (DNMT3A) mRNA"  
BASE COUNT 571 a 505 c 575 g 540 t  
ORIGIN

Alignment Scores:  
Pred. No.: 8.3e-100 Length: 2191  
Score: 2412.00 Matches: 440  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 99.55% Mismatches: 0  
Query Match: 49.13% Indels: 0  
DB: 9 Gaps: 0  
US-09-720-086-5 (1-908) x AK025230 (1-2191)  
QY 467 IleAspGluArgThrArgGluValTyrGluValArgGlnLysCysArgAsnIle 486  
DB 2 ATTGATGAGGCGCACAAGAGCGGCTGTGTAGAGATGGGCGAGAAGTCCGGAACATT 61  
QY 487 GluAspIleCysIleSerCysGlySerIleAsnValThrLeuGlnHisProLeuPheIle 506  
DB 62 GAGGACATCTGCATCTCTCTGTGGAGCTTCATATGTTACCTGGAAACACCCCTTTCGTT 121  
QY 507 GlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAsp 526  
DB 122 GGAGGATGTGCCAATAATGCAAGAACTGCTTCTGAGTGTGCTACACGATCGACGAC 181  
QY 527 AspGlyTyrGlnSerTyrCysThrIleCysGlyGlyArgGluValLeuMetCysGly 546  
DB 182 GACGGCTACACGATCTTACATGACCATCTGCTGTGGGGCGGTGAGGTGCTCATGTGGGA 241  
QY 547 AsnAsnAsnCysCysArgCysPheCysValGluCysValAspLeuValGlyProGly 566  
DB 242 AACCAACATCTGTCGAGGTCTTTTGGTGAGTGTGAGACTCTTGTGTGGGGCCGGG 301  
QY 567 AlAlaGlnAlaAlaIleLysGluAspProTyrAsnCysTyrMetCysGlyHisGly 586  
DB 302 GTGGCCAGGACCACTTAAGAAAGACCCTGTGAATGCTGTAATGTGGGGCACAAGGCT 361  
QY 587 ThrTyrGlyLeuLeuArgArgGluAspTyrProSerArgLeuGlnMetPhePheAla 606  
DB 362 ACCTAGGGGTGTGGGGCGGCGCAGAGAGATGAGCTCCCGGCTCCAGATGTTCTTGCT 421  
QY 607 AsnAsnHisAspGlnGluPheAspProLysValTyrProProAlaGluLys 626  
DB 422 AATTAACACAGACAGGAATTTGACCTCCAAAGCTTAACTCCAGCTGACGAGAG 481  
QY 627 ArgLysProIleArgValLeuSerLeuPheAspGlyIleAlaThrGlyLeuValLeu 646  
DB 482 AGAAGCCATCGGGGGGTGCTCTCTTGTGATGATGCTACAGGGCTCTGTGCTG 541  
QY 647 LysAspLeuGlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIle 666  
DB 542 AAGGACTTGGGCAATTCAGGTGGACCGCTACATTGCTCGGAGGTGTGTGAGACTCCATC 601  
QY 667 ThrValGlyMetValArgHisGlnIleGlnIleLysIleMetTyrValGluValAspVal 686  
DB 602 AGGTGGGATGTGTGGCGCCAGGAGAGATATGATCTGGGGCGTCCGAGGCTC 661  
QY 687 ThrGlnLysHisIleGlnIleTyrPheProPheAspLeuValIleGlyGlySerProCys 706  
DB 662 ACACAGAGCATATCCAGAGATGGGGCCATTGCATCTGTGATTTGGGGGCAAGCTCCGTC 721  
QY 707 AsnAspLeuSerIleValAsnProAlaArgLysGlyLeuTyrGlnIleThrGlyArgLeu 726  
DB 722 AATGACCTCTCCATCGTCAACCTGCTCGCAGAGGCTCTTACACAGGCGACCTGCGGCTC 781  
QY 727 PhePheGluPheTyrArgLeuLeuHisAspAlaArgProLysGlnIleGlyAspAspArgPro 746  
DB 782 TTTCTTGTAGTTCTACCGCTCTCTGTCATGATGCGGGCCAGAGGGAGATGATCGCCC 841  
QY 747 PhePheTyrLeuPheGluAsnValAlaAlaMetGlyValSerAspLysArgAspIleSer 766  
DB 842 TTTCTTGGCTCTTTGAGATGTGTGGCATGTGGGCTTATGTACAAAGAGGACATCTCG 901  
QY 767 ArgPheLeuGlnSerAsnProValMetIleAspAlaLysGluValSerAlaAlaHisArg 786

```

Db 902 CGATTTCGAGTCCAAACCTGTGTATGATGCGCAAGAGAGTGTGACGTGCACACAGG 961
Qy 787 AlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArgProLeuAlaSerThrValAsn 806
Db 962 GCCCGCTACTTCTGGGGTAACCTTCCGGTATGAACAGGCCGTTGGCATCCACTGTGAAT 1021
Qy 807 AspLysLeuGluLeuGlnGlyCysLeuGluHisGlyArgIleAlaLysPheSerLysVal 826
Db 1022 GATAAGCTGGAGCTGCAGGAGTGTCTGGAGCATGCGAGGATAGCCAAAGTTCAGCAACGTG 1081
Qy 827 ArgThrIleThrThrArgSerAsnSerIleLysGlnGlyAspGlnHisPheProVal 846
Db 1082 AGACCACTTACTACGAGGTCAACATCCATAAGCAGGCGCAAGCAGCACTTTCTCTGC 1141
Qy 847 PheMetAsnGluLysGluAspIleLeuTrpCysThrGluMetGluArgValPheGlyPhe 866
Db 1142 TTCATGAATGAGAAAGAGGACATCTTATGGTGCACATGAATGGAAGGGTATTTGGTTTC 1201
Qy 867 ProValHisTyrThrAspValSerAsnMetSerArgLeuAlaArgGlnArgLeuGly 886
Db 1202 CCAGTCCACTATATCTGACGTCTCCAAACATGAGCCGCTTGGCAGGAGAGACTGTCTGGC 1261
Qy 887 ArgSerTrpSerValProValIleArgHisLeuPheAlaProLeuLysGluTyrPheAla 906
Db 1262 CGGTCAATGAGCGTGCAGTCAATCGCCACCTCTTCGCTCCGCTGAAGGAGTATTTGCG 1321
Qy 907 CysVal 908
Db 1322 TGTGTG 1327

RESULT 7
LOCUS AC009474 168651 bp DNA linear PRI 08-NOV-2000
DEFINITION Homo sapiens BAC clone RP11-179G23 from 2, complete sequence.
ACCESSION AC009474
VERSION AC009474.4 GI:8748849
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 168651)
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
MEDLINE
9847074
PUBMED
REFERENCE 2 (bases 1 to 168651)
Molvane,E., Stoneking,T., Laplant,Y. and McBill,B.
The sequence of Homo sapiens BAC clone RP11-179G23
JOURNAL Unpublished
3 (bases 1 to 168651)
Waterston,R.H.
Direct Submission
Submitted (24-AUG-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 168651)
Waterston,R.H.
Direct Submission
Submitted (27-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 168651)
Waterston,R.
Direct Submission
Submitted (08-NOV-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jun 27, 2000 this sequence version replaced gi:7630880.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc

```

Contact: sapiens@watson.wustl.edu  
----- Summary Statistics  
Center project name: H\_NH0179G23  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Oosagawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) VECTOR: pBAC1.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-547F18; the clone sequenced to the right is RP11-444B4, 200 base pair overlap. Actual end of this clone is at base position 1871 of RP11-444B4.

#### FEATURES

Location/Qualifiers	Source
1..168651	/organism="Homo sapiens"
/db_xref="taxon:9606"	
/chromosome="2"	
/map="2"	
/clone="RP11-179G23"	
/clone_lib="RPC1-11"	
13..497	/rpt_family="L2"
616..683	/rpt_region
1615..1815	/rpt_family="L2"
2577..2748	/rpt_family="MIR"
2954..3144	/rpt_family="MIR"
3149..3465	/rpt_family="MER1_type"
3510..3567	/rpt_family="L1"
3569..3760	/rpt_family="MER1_type"
3827..4198	/rpt_family="MER1_type"
5656..6129	/rpt_family="MER1_type"
6205..6316	/rpt_family="MalR"
6428..6686	/rpt_family="L2"
7247..7437	/rpt_family="L1"
7819..7868	/rpt_family="MER53"





## FEATURES

source	gene	misc_feature	CDS
1..4135	1..4135	1..4135	269..2788
/organism="Mus musculus"	/organism="Mus musculus"	/organism="Mus musculus"	
/db_xref="taxon:10090"	/db_xref="taxon:10090"	/db_xref="taxon:10090"	
/chromosome="2"	/chromosome="2"	/chromosome="2"	
/map="84.0 cM"	/map="84.0 cM"	/map="84.0 cM"	
1..4135	1..4135	1..4135	
/gene="Dnmt3b"	/gene="Dnmt3b"	/gene="Dnmt3b"	
1..4135	1..4135	1..4135	
/gene="Dnmt3b"	/gene="Dnmt3b"	/gene="Dnmt3b"	
/note="Similar to E2T sequences deposited in GenBank	/note="Similar to E2T sequences deposited in GenBank	/note="Similar to E2T sequences deposited in GenBank	
Accession Numbers A116694, A119979, A117277, A101068	Accession Numbers A116694, A119979, A117277, A101068	Accession Numbers A116694, A119979, A117277, A101068	
AA07106, and AA575617"	AA07106, and AA575617"	AA07106, and AA575617"	
269..2788	269..2788	269..2788	

BASE COUNT	1043	a	1048	c	1106	g	938	e
ORIGIN								

Alignment Scores:	
Pred. No.:	1.55e-90
Score:	2311.00
Percent Similarity:	61.22%
Biological Similarity:	46.95%
Match:	45.04%
DB:	10
	Gaps:
	15

US-09-720-086-5 (1-908) x AF068627 (1-4135)

Qy	4	SerGlyProGlyAspThrSerSer	-----SerLeuGluArgGluAsp	18
Db	22	AGCGCCCAAGTAAACGTACGCGACACCATCGGCGCCGGAGATTGGCGAACCCGACACTCC		81
Qy	19	AspArgIysGlnGlyGlnGlnGlnGlnGlnGlnGly-LysGlnGluArgGlnGln		38
Db	82	GCGCAGCCCGCGGCGACAGCCCGGCGCGATCGCGCGCGCGCGGTACAGCGCTCA		141
Qy	38	oSerAlaThrAlaArgIysValGlyArgProGlyArgIysArgIysHisPro	---ProVa	57
Db	142	CGACAGGCCCGCTGAGCGTTGTGCAGACCTTGAGAACTCAGGATATATACCTTCCAGA		201
Qy	57	IglUserSerAspThrProIysAspProAlaValThrThrIysSerGlnProMetAlaG		77
Db	202	CGCGGAGTCTCCCTCCGCCCATTCATAGTGCCTTGGACCAAAATCCAGGCGCTTCTTTCA		261
Qy	77	naSPserGlyProSerAsp	-----	83
Db	262	GGAAACATGAAAGGAGACAGACGACATCTGATAGAAAGAGAGGGTGCACAGGGGTATGA		322
Qy	84	-----LeuLeuProAsnGlyAspLeuGlnIysArgSerGluProGlnProGlnGlnGln		101

Db	322	GGAGTGCATTATGTTAAATGGAACTTCAGTCAGCAGCTCTCAGACACGAAGAGTCTCC	381
Qy	101	yserProAlaAlaGlyGlnIysGlyValAlaProAlaGlnGlyGlnGlyThrGlnThrPr	121
Db	382	CTCACCC-----CC	390
Qy	121	oProGluAlaSerArgAlaValGlnIuSnGlyCysCysValThrIysGlnGlyArgGlyAl	141
Db	391	AGTTTGAGAGCAATCTGCACAGAGCCAGCTTGCACACACAGAGACCAAGGCCGAGGTC	450
Qy	141	aSerAla-----	143
Db	451	AAGCTCCGGGCTGCTTAAGAGGAGGCTTCAGGCTTCGAATTACAGCAGACATATAC	510
Qy	144	-GlyGlnGly-----	146
Db	511	AGGAGATGGAGAGACAGATGATGAATAGATGAGGATGGCTCTGATATTCTAATCC	570
Qy	147	-----LysGlnGlnIysGlnIhrzsnIleGlnSerMetIysMetGlnGlySerzAr	163
Db	571	AAAGCTCACCCGCTGAGACCAAGGACCAAGACCGGCTT-----GAAAGCCGGC	621
Qy	163	gGlyArgLeuArgGlyGlyIysGlnIYrpbIuSerSerLeuArgIlnArpPromePror	183
Db	622	TGTCGGAACCCGACATATGACATGGAGACCTCCAGCTTGGAGAGGCAAAAGCTCCCCCAG	681
Qy	183	gLeuThrPheGlnAlaGlyAspProTyTrIleSerIysArgIysArgSpGluTIPLE	203
Db	682	AATCACCCGAGGTGGGAGGAGCGGCACCATG-----CAGAGTACCC	726
Qy	203	uAlaArgTrpIysArgGluAlaGlnIysIysAlaIysValIleAlaValMetAsnAlaIva	223
Db	727	TGNGAGTTTCCGAGTACACAGGCTTCGGAGAGCGTGCAGCATGCTTCAGCAAGACGCC	786
Qy	223	IglIuAsnGlnAlaSerGlyIysGlnSerGlnIysValIgluAlaIserProProAlaIva	243
Db	787	ATGTGATCTCCCTGCGCACGCTGCATCTC-----ATGAAAGAGTACACCTTAAGAGCT	840
Qy	243	IglInIuProThrAspProAlaSerProThrValAlaThrThrProGluProAlaIglYl	263
Db	841	CAGTACCCCA-----TCAGTTGACTTGAGCCAGATGGAGATCAGAGAGGTAT	888
Qy	263	yAspAlaGlyAspIysAsnAlaThrIysAlaAlaAspAspGluProGluIYrGluAspGl	283
Db	889	GGATACACACACAGGTGATGTCAGAGAGCAGAGATGAGACAGACACAGATACAGATGA	948
Qy	283	yArgGlyPheGlyTrIleGlyGlnIuValTrpGlyIysLeuArgIysPheSerTrpTrPr	303
Db	949	TAAAGATTGGAAATAGTGAACCTCGTGTGGGAAAGATCAAGGGCTTCTCTGTGTGCC	1008
Qy	303	oGlyArgIleValSerTrpTrpMetThrGlnArgSerArgAlaAlaGlnGlyThrArgTr	323
Db	1009	TGCCATGTGTGTCTCTGGAAGCCACCTCCAAAGCAGACAGCCATGCCCGGAATGCGCTG	1068
Qy	323	pValMetTrpPheGlyAspGlyIysPheSerValValCysAlaGlnIysLeuMetProIe	343
Db	1069	GGTACAGTGGTTGGTGATGGCAAGTTTCTGTGAGATCTCTGTGCAAACTGTGTGCTCT	1128
Qy	343	uSerSerPheCysSerAlaPheHisGlnAlaThrTyTrzsnIysGlnProMetTyArgIly	363
Db	1129	GGGGCTGTTACGCACGACCTTAATGTGCTACCTTCATTAAGCTGGTTCTTATAGAA	1188
Qy	363	saIleIleTyGluValIuGlnIuValIaSerSerArgAlaGlnIysLeuPheProAlaCy	383
Db	1189	GGCCATGTACACACTGTGGAGAAAGCCAGGGTTGCAAGCTGGCAAGACCTTC-----	1240
Qy	383	hHisAspSerAspIuSerAspSerGlyIysAlaIvalaGluAlaGlnIysGlnMetI1	403
Db	1241	-----TCCAGCAGTCTCTGGAGAGTCACTGGAGGCCACGCTGAAGCCATCT	1287
Qy	403	eGluTrpAlaIuGlnGlyGlyPheGlnProSerGlyProIysGlyLeuGluTrpProGlnGl	423
Db	1288	GGATGTGGCCACAGGTGGCTTCAAGCTTCTGTGGATTCGAGGGCTTCAAAACCAACAGAA	1347

QY 423 uGlu---LysAsnProTyrLysGluValTyrThrAspMetTrpValGluProGluAlaA 442  
Db 1348 GCAACAGAGAACAAAGTGAAGAGCGACACCAAT-----GACTTGC 1392  
QY 442 aAlaTyrAlaProProPProAlaLysLysLysProArgLysSerThrThrGluLysProLY 462  
Db 1393 TGCTTCTGAGTCCCTCCACCCCAAGCGCTCAAGACAAATAGCTATGGCGGGAAGACCG 1452  
QY 462 sValLysGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 482  
Db 1453 AGGGAG-----GATGAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1503  
QY 482 sCysArgAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 502  
Db 1504 CAAGGCAATCTCGAAGAGCGCTGTTGCTGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1563  
QY 502 sProLeuPheLysGlyMetCysGlnAsnCysLysAsnCysPheLeuGluCysAlaTY 522  
Db 1564 CCCCCTCTTTGAGGGTGGGCTCTGTCAAGATTGGCGGATCGCTTCTTAGAGCTCTTCTA 1623  
QY 522 rGlnTyrAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyGlyArgGluVa 542  
Db 1624 CATGTATGATGAGAGCGGTATCACTGCTCTGACCGTGTGCTGAGGCGCGTGAAC 1683  
QY 542 lLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysValAspLeuLe 562  
Db 1684 GCTCTGTGAGTAAACACAGCTGTGTCAGATGCTTCTGTGGAGTGTCTGGAGGTCT 1743  
QY 562 uValGlyProGlyValAlaGlnAlaLysLysLysLysLysLysLysLysLysLysLys 582  
Db 1744 GGTGGCGCAGGACAGCTGAGATGCCAAGCTGCAGGACCTCGAGCTGTATATGTG 1803  
QY 582 sGlyHisGlyGlyTyrGlyLeuLeuArgArgGluAspTyrProSerArgLeuGL 602  
Db 1804 CTCTCTCAGCGTCCGATGGGTCTCCGAGCAGAGAGAGATTGGAACATCGCGCTCA 1863  
QY 602 mMetPhePheAlaAsnAsnHisAsp---GlnGluPheAspProProLYsValTyrProPr 621  
Db 1864 AGACTTCTTCACTACTGATCTGAGCTGGAAGATTTGAGCCACCAAGATTGATCCACAG 1923  
QY 621 ovalProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIleAlaTh 641  
Db 1924 AATTCTGCGACGCAAAAGAGAGCCCATTAGAGTCTCTCTGTTTGTATGGGAATTGCAAC 1983  
QY 641 rGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSerGluVa 661  
Db 1984 GGGGTACTTGGTCTCAAGAGTTGGGTATTAAAGTGAAGATGACATTGCTCCGAAAGT 2043  
QY 661 lCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyrValG1 681  
Db 2044 CTGTGCAGAGTCCATCGCTGTGGAACTGTTAAGCATGAAGGCCAGATCAAAATATGTCAA 2103  
QY 681 YAspValArgSerValThrGlnLysHisIleGlnGluTyrGlyProPheAspLeuVal11 701  
Db 2104 TGAGCTCGGAAATCACCAGNAAATATTGAAGAGTGGGCGCCCTTCGACTTGGTAT 2163  
QY 701 eGlyGlySerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeuTyrGL 721  
Db 2164 TGGTGAAGGCCCATGAATGATCTCTTAAGTCATCTCCCGCAAGAGTTATATGA 2223  
QY 721 uGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgProLYsGL 741  
Db 2224 GGGCACAGGAAGGCTCTTCTCGAGTTTACCCTTCTGAAATATACCGCCGCCCAAGGA 2283  
QY 741 uGlyAspAspArgProPhePheThrLeuPheGluAsnValValAlaMetGlyValSerAs 761  
Db 2284 GGGCGCAACACCGTCCATCTTCTGGATGTTCCAGAAATGTTGTGGCCATGAAGATGAATGA 2343  
QY 761 pLysArgAspIleSerArgPheLeuGluLysSerAsnProValMetIleAspAlaLysGluVa 781  
Db 2344 CAAGAAAGACATCTCAAGATTCTGGCATGTAAACCCAGTATGATCGATCGATCAAGT 2403

QY 781 lSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArgProLe 801  
Db 2404 GTCTGCTGTACAGGGCCCGTACTTCTGGGTAACTTACCGGAATGAACAGGCCGT 2463  
QY 801 uAlaSerThrValAsnAspLysLeuGluCysLeuGluHisGlyArgIleAl 821  
Db 2464 GATGGCTTCAAGAATGATAGCTCGAGCTGCAGGACTGCCTGGAGTTTCAGTAGGACAGC 2523  
QY 821 aLysPheSerLysValArgThrIleThrArgSerAsnSerIleLysGlnGlyLysAs 841  
Db 2524 AAAGTTAAAGAANGTCAGACATTAACCAAGTCGAACCTCCATCAGACAGGCAAAAA 2583  
QY 841 pGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTyrCysThrGluMetG1 861  
Db 2584 CCAGCTTTTCCCTGTAGTCATGAATGCAAGGACGACGTTTGTGTGCTGCACTGAGCTCGA 2643  
QY 861 uArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeuAlaAr 881  
Db 2644 AAGATCTTCGGCTTCCCTGCTCTACACAGGACGTGTCAACATGGCGCCGCGCCG 2703  
QY 881 gGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAlaProLe 901  
Db 2704 TCAGAAAGCTGTGGCGAGTCTCTGGAGTGTACCGGTATCAGACACCTGTTTGGCCCTT 2763  
QY 901 uLysGluTyrPheAlaCys 907  
Db 2764 GAAGGACTACTTTCCTGT 2782  
RESULT 9  
AF151970 4278 bp mRNA linear ROD 08-JUN-2000  
LOCUS Mus musculus DNA cytosine-specific methyltransferase isoform 2  
DEFINITION (Dnmt3b) mRNA, complete cds.  
ACCESSION AF151970  
VERSION AF151970.1 GI:8347119  
KEYWORDS Mus musculus.  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
REFERENCE 1 (bases 1 to 4278)  
AUTHORS Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and Shen,Y.  
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing isoforms in mouse embryonic tissue  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4278)  
AUTHORS Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and Shen,Y.  
TITLE Direct Submission  
JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular Biology, Institution of Basic Medical Sciences, Chinese Academy of Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005, P.R.China  
FEATURES  
Location/Qualifiers  
1..4278  
/organism="Mus musculus"  
/strain="KM"  
/db\_xref="taxon:10090"  
/dev\_stage="8-9 day old embryo"  
1..4278  
/gene="Dnmt3b"  
412..2931  
/gene="Dnmt3b"  
/function="methylates cytosine in DNA"  
/note="alternatively spliced"  
/codon\_start=1  
/product="DNA cytosine-specific methyltransferase isoform 2"  
/protein\_id="AAF74516.1"  
/db\_xref="GI:8347120"  
/translation="MKGDSRHLNEEGASGYEEICIVNGNFSQSSDTRKADAPSPVLE  
AICTPEVCTPETRGRSSRLSKREVSSLLNYTQDMTGDRDRDEVDGNGSDILMPK





```
OY 542 lLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysValAspLeuLe 562
Db 1827 GCTGCTGTGCAGTAACACAAAGCTGCTGCAGATGCTTCTGTGTGAGATGCTGCAGGAGTCT 1886

OY 562 uValGlyProGlyValAlaGlnAlaAlaIleGlyGluAspProTyrAsnCysTyrMetCys 582
Db 1887 GGTGGCGCAGGCACAGCTGAGGATGCCAAGCTGCAGGAACCTGCAGGCTGTATATGTG 1946

OY 582 sGlyHisGlyGlyThrTyrCysLeuLeuArgArgGluAspTyrProSerArgLeuGly 602
Db 1947 CTTCCCTCAGCGCTGCATGGGTCTCTCCAGCGAGAGATGTGGACATCGCGCTGCA 2006

OY 602 nMetPhePheAlaAsnAsnHisAsp---GlnGluPheAspProProGlyValValTyrProPr 621
Db 2007 AGACTTCTTCACTACTGCTGACCTCGAAGAAATTTGAGCCACCAAGTGTGACCCAGC 2066

OY 621 oValProAlaGlyArgLeuProIleArgValLeuSerLeuPheAspGlyIleAlaIth 641
Db 2067 AATTCTTCGACGCAAAAGAGAGCCCAATTAGAGTCTCTCTCTGTTTGTAGTGAATTCGAAC 2126

OY 641 rGlyLeuLeuValLeuAspLeuGlyIleGlnValAspArgTyrIleAlaSerGluVal 661
Db 2127 GGGGTACTTGTGCTCAAGGAGTGGGTATTAAAGTGGAAAGTACATTGCTCCTCGAAGT 2186

OY 661 lCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLeuMetTyrValGly 681
Db 2187 CTGTGCAGATCCATCGCTGTGGGAATGTTAAGCATGAAGCCAGCATCAATATGTCAA 2246

OY 681 yAspValArgSerValThrGlnLysHisIleGlnGluTyrGlyProPheAspLeuValIle 701
Db 2247 TGAGTCTCGGAAATCACCAGAAATATTGAAGATGGGGCCCTTCGATCTGTGAT 2306

OY 701 eGlyGlySerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeuTyrGly 721
Db 2307 TGGTGGGAAGCCCATGCAATGATCTCTTAACGTCAATCTCGCCGCAAAAGGTTTATATGA 2366

OY 721 uGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgProLysGly 741
Db 2367 GGGCACAGGAAGGCTCTTCTTCGAGTTTACCACTTCTCGAATATATACCCGCCGCCCAAGGA 2426

OY 741 uGlyAspAspArgProPheThrLeuPheGluAsnValValAlaMetGlyValSerAs 761
Db 2427 GGGCGCAACCGTCCATCTTCTGATGTTTCGAGAATGTTGTGGCCATGAAAGTGAATGA 2486

OY 761 pLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLysGluVal 781
Db 2487 CAAGAAAGACATCTCAAGATCTCTGGCATGTAACCCAGTGCATGCATGCATCAAGGT 2546

OY 781 lSerAlaAlaHisArgAlaArgTyrPheThrGlyAsnLeuProGlyMetAsnArgProLe 801
Db 2547 GTCTGCTGCTCAGAGGCCCGGTACTTCTGGGTAACTTACCTACCCGGAATGAACAGCCCGT 2606

OY 801 uAlaSerThrValAsnAspLeuLeuGlnGluCysLeuGluHisGlyArgIleAl 821
Db 2607 GATGCTTCAAGAAGATGATAGCTCGAGCTCGAGGATGCTGAGGATTCAGTAGGACAGC 2666

OY 821 aLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGlyLysAs 841
Db 2667 AAAGTTAAAGAAAGTCAGACAAATACCAACCAAGTCGAATCTCATCAGACAGGCAAAA 2726

OY 841 pGlnHisPheProValPheMetAsnGlnLysGluAspIleLeuTyrCysThrGluMetGly 861
Db 2727 CCAGCTTTTCCCTGTAGTCAATGATGCGCAAGGACGAGTGTGCTGCTGCTGCTGCTG 2786

OY 861 uArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgGluAlaAr 881
Db 2787 AAGGATCTTCCGCTTCCCTGTCTCACTACACGAGCAGTGTCTCAACATCGGGCCGCGCCG 2846

OY 881 gGlnArgLeuLeuGlyArgSerTyrPheValProValIleArgHisLeuPheAlaProLe 901
Db 2847 TCAGAGCTGCTGGGCGAGTCTCTGGAGTGTTACCGGTATCATCAGACACCTGTTTGGCCCTT 2906

OY 901 uLysGluTyrPheAlaCys 907
```

```
Db 2907 GAAGGACTACTTTCCTGT 2925

RESULT 10
LOCUS AF068626 4195 bp mRNA linear ROD 06-DEC-1999
DEFINITION Mus musculus DNA cytosine-5 methyltransferase 3B1 (Dnmt3b) mRNA,
alternatively spliced, complete cds.
ACCESSION AF068626
VERSION AF068626.2 GI:6449469
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4195)
AUTHORS Okano,M., Xie,S. and Li,E.
TITLE Cloning and characterization of a family of novel mammalian DNA
(cytosine-5) methyltransferases
JOURNAL Nat. Genet. 19 (3), 219-220 (1998)
MEDLINE 98324766
PUBMED 9662389
REFERENCE 2 (bases 1 to 4195)
AUTHORS Xie,S., Okano,M. and Li,E.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,
Charlestown, MA 02129, USA
REFERENCE 3 (bases 1 to 4195)
AUTHORS Okano,M., Chijiwa,T., Sasaki,H. and Li,E.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,
Charlestown, MA 02129, USA
REMARK Sequence update by submitter
COMMENT On Nov 18, 1999 this sequence version replaced gi:3327979.
FEATURES
Location/Qualifiers
1..4195
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosomes="2"
/map="84.0 cm"
1..4195
/gene="Dnmt3b"
1..4195
/gene="Dnmt3b"
/notes="similar to EST sequences deposited in GenBank
Accession Numbers AA116694, AA119979, AA177277, AA210568,
AA407106, and AA575617"
269..2848
/gene="Dnmt3b"
/function="de novo DNA methylation"
/notes="alternatively spliced product; contains Cys-rich
region; C-terminal region is similar to corresponding
region of DNA cytosine-5 methyltransferase 3A"
/codon_start=1
/product="DNA cytosine-5 methyltransferase 3B1"
/db_xref="GI:6449470"
/translat="MKGDSRHLNEEGASGYECIIIVNGNFSQSDSTKADAPSPVLE
AITEPVCTPSTRSSRLSKREVSLLNYTQDMTGDGDRDDVDGDSGLMPK
LRTETKDTTRSEPAVRTHSNGTSSLRQASPRITRQGRHHVQVEYVPEPATR
SRRRASSASPTWSPASVDFMEETPKSVTPSDLSQDQDQDMTTOVDAESRD
GDSTEQDDKEFGIDGLVMGKIKGFSWPMVVMVSKATSKRAMPCMRVOWFGDGK
SETSDKLVALGLFOHFNLATFNKLVSRKAMHTLEKARVACKTFFSSPGLSD
OLKPMWAHGGFKPTGIEGLKPNKOPVNVKSKVRNRLEPRRNRKRRRTTN
DSAAESPPKLTNSYGGKDRGEDESERMASVNTNNKNDLRCCLSCGKNPV
PFLPFGGLCQCRDRPLELPMYDQYQCTVCCGRELILLCSNCCRCFCVEC
LEVLVAGTAEDAKLQEPWSCMCLPORCHGLRRKDWMLQDFTTDPDLSEFER
PKLYPAIPAAKRRPIRLSLSDGATGLYLKELGKVEKYTASEVCAESVAGTVKH
EGKYKVNDRKTKTKEWGFPLVIGGSPCNDLSNVNPAKGLYEGTGLRLEPBY
HLANTYRPEKGNRPFPKMFNVVMKVNPKKDI SFLACNPMIDA IKVSAHARY
PMGNLCKMNRPMVASKDKLELQCLFSTAKLKKVQITTKNSIKGKQLFPVV
MNGKDDVLWCTELERIFQPPAHYTDVSMGRGARQKLLGRSWSVPVIRHLFAPLKDIF
ACE"
```

BASE COUNT 1061 a 1059 c 1125 g 950 t  
ORIGIN

## Alignment Scores:

Pred. No.:	2,62e-90	Length:	4195
Score:	2206.00	Matches:	455
Percent Similarity:	59.78%	Conservative:	135
Best Local Similarity:	46.10%	Mismatches:	268
Query Match:	44.94%	Indels:	129
DB:	10	Gaps:	15

US-09-720-086-5 (1-908) x AF068626 (1-4195)

```
QY 4 SerGlyProGlyAspThrSerSerSer-----SerLeuGluArgGluAsp 18
Db 22 AGCGGCCCAAGTAAACGTACGCGAGCATCGCGCCGAGATTCTGCACCCGACACTCC 81
QY 19 AspArgLysGluGlyGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 38
Db 82 GCGCCGCCCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCA 141
QY 38 oSerAlaThrAlaArgLysValGlyArgProGlyArgLysArgLysHisPro---ProVa 57
Db 142 CGACAGGCGCGCGCTGAGGCTGTGCGACACCTTGGAACCTCAGGTATATACCTTTCCGA 201
QY 57 LgluSerSerAspThrProLysAspProAlaValThrThrLysSerGlnProMetAlaG1 77
Db 202 CGCGGATCTCCCTCCCTCCCATCATAGTGTGCGGAGCCAAATCCAGGGCTTTCTTCA 261
QY 77 nAspSerGlyProSerAsp----- 83
Db 262 GGAACCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 321
QY 84 -----LeuLeuProkengLysAspLeuGluLysArgSerGlnProGlnProGluGluG1 101
Db 322 GAGTGCATTATGTTATATGGAATTCACAGTACAGTCTTCAGACACGAGGAGTGCTCC 381
QY 101 ySerProAlaAlaGlyGluLysGlyGlyAlaProAlaGluGlyGluGlyThrGluThrPr 121
Db 382 CTCACCC-----CC 390
QY 121 oProGluAlaSerArgAlaValGluAsnGlyCysCysValThrLysGluGlyArgGlyAl 141
Db 391 AGCTTGAGGACATCTGCACAGGCGAGCTGTGCACACACAGACACGAGAGCGCAGGTC 450
QY 141 aserAla----- 143
Db 451 AAGCTCCGCGCTGTCTAAGAGGAGGCTCCAGCCTTCTGAATTACGACGACGATGAC 510
QY 144 -GlyGluGly----- 146
Db 511 AGGAGATGAGGAGACAGATGATGAAGATGATGATGATGATGATGATGATGATGATGAT 570
QY 147 -----LysGluGluLysGluLysGluLysGluLysGluLysGluLysGluLysGlu 163
Db 571 AAGGCTACCCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 621
QY 163 gGlyArgLysArgGlyGlyLysGluGlyLysGluGlyLysGluGlyLysGluGlyLysGlu 183
Db 622 TGTCGAGACCGGACATGACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 681
QY 183 gLeuThrPheGluAlaGlyLysAspProLysArgLysArgLysArgLysArgLysArgLys 203
Db 682 AATCACCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 726
QY 203 uAlaArgTrpLysArgGlyAlaGluLysLysAlaLysValIleAlaValMetAsnAlaVa 223
Db 727 TGTGAGATTCTCGGCTCAACAGGTCGAGAGCGTGCAGACATCGCTTCAGAGACAGCC 786
QY 223 LgluGluAsnGluAlaSerGlyGluSerGlnLysValGluGluAlaSerProProAlaVa 243
Db 787 ATGTCATCCCTCCGCGAGCGTCGACTTC-----ATGGAAGAGTGAACCTTAAGAGCGT 840
```

```
QY 243 LgluGlnProThrAspProAlaSerProThrValAlaThrThrProGluProValGlyG1 263
Db 841 CAGTACCCTCA-----TCAGTTGACTTGACCGACAGATGATGACGAGGGTAT 888
QY 263 yAspAlaGlyAspLysAlaThrLysAlaAlaAspAspGluProGluLysGluAspG1 283
Db 889 GATATACCACACAGTGTGATGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 948
QY 283 yArgGlyPheGlyIleGlyGluLeuValTrpGlyLysLeuArgGlyPheSerTrpTrpPr 303
Db 949 TAAAGATTGGAAATAGGTACCTCTGTGGGAAAAAGATCAAGGCTTCTCTGCTGGGCC 1008
QY 303 oGlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGluGlyThrArgTr 323
Db 1009 TGGCAGTGTGTGTCTCTGGAAAGCCACCTCCAGACGACAGGAGGAGGAGGAGGAGGAG 1068
QY 323 pValMetTrpPheGlyLysAspGlyLysPheSerValValCysValGluLysLeuMetProLe 343
Db 1069 GTATACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1128
QY 343 uSerSerPheCysSerAlaPheHisGluAlaThrTyraAsnLysGlnProMetTyraArgLy 363
Db 1129 GGGGCTGTTCAGCCAGCATTTAATCTGCTTCAATACCTGTGTTCTTATAGGAA 1188
QY 363 sAlaIleTyrgLysValLeuGluAlaAlaSerSerArgAlaGlyLysLeuPheProAlaCy 383
Db 1189 GGCATGTATCACACATCTGTGAGAAAGCCAGGGTTCGAGCTGGCAAGACCTTC----- 1240
QY 383 sHisAspSerAspGluSerAspSerGlyLysAlaValGluValGluAsnLysGlnMetI1 403
Db 1241 -----TCCAGCGATCTGTGAGAGTCACTGAGGAGCACCTGAAAGCCATGCT 1287
QY 403 eGluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGluPro----- 420
Db 1288 GAGTGGGCCACGATGTGCTTCAAGCTTACTGGAGTCAAGGGCTCAAAACCAAGAA 1347
QY 421 -----ProG1 422
Db 1348 GCAACCAAGTGTTAATATAGTCGAAGGTGGCTGCTTCAGACATGAGAACTAGAACCCAG 1407
QY 422 uGluGluLysAsnProTyrgLysGluValTyThrAspMetTrpValGluProGluAla1 442
Db 1408 GAGACCGCGAAGAAAGTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1452
QY 442 aAlaTyraLysProProProAlaLysLysProAlaGlySerThrThrGluLysProLy 462
Db 1453 TGCTTGTGAGTCCCCCCCCACCAAGCGCTCAAGACAAATATATGCGGGAGGAGCG 1512
QY 462 sValLysGluIleIleAspGluArgThrArgGluArgLeuValTyrgLysValArgLys 482
Db 1513 AGGGGAG-----GATGAGGAGAGCCGAGAACCGATGCTTCGAACTCAACAA 1563
QY 482 sCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThrLeuGluH1 502
Db 1564 CAAGGCGCAATCTGGAAGACCGCTGTTGTCTGTGAAAGAAAGAACCCGTGTCTTCCA 1623
QY 502 sProLeuPheIleGlyGlyMetCysGluAsnGlyLysAsnGlyPheLeuGluCysAlaTy 522
Db 1624 CCCCCTTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1683
QY 522 rGlnTyraAspAspAspGlyTyrgLysSerTyrgLysThrIleCysGlyGlyValArgGluVa 542
Db 1684 CATGTATGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1743
QY 542 LLeuMetCysGlyAsnAsnAsnGlyCysArgCysPheCysValGluCysValAspLeuLe 562
Db 1744 GCTGCTGTGCAATAACAAGGCTGTGAGATGCTTCTGTGTGAGTGTGTGAGGAGGAGG 1803
QY 562 uValGlyProGlyAlaAlaGluAlaIleLysGluAspProTrpAsnGlyTyrgMetCy 582
Db 1804 GGTGGGCGAGGCGACAGCTGAAGCTCAAGAGCTGCAAGAACTTGAAGCTGTATATGTG 1863
QY 582 sGlyHisLysGlyThrTyrgLysLeuLeuArgArgGluAspTrpProSerArgLeuG1 602
```

```
Db 1864 CCTCCCTCAGCGCTGGGTCCTCCGACGACGAGAAAGATTGGAAACATCGCGCTGCA 1923
Qy 602 wmerPhePheAaAsnHisAasp---GloclupheAspProProlyValTyProPr 621
Db 1924 AGACTTCTTACTACTGATCTGACCTGGAAGAATTGAGCCACCCCAAGTTGATCCAGC 1983
Qy 621 ovalProAlaGluLysArgLysProLleArgValLleuSerLeuPheAspGlyLleAlaTh 641
Db 1984 AATTCTCTGACGACCAAGAGGCCCATAGAGTCTGCTCTGTTGATGAGGATGCAAC 2043
Qy 641 rGlyLeuLeuValLeuLysAspLeuGlyLleGlnValAspArgTyrlleAlaSerGluVa 661
Db 2044 GGGGTACTTGTGTCGCAAGAGTGGGTATTAAAGTGGAAAGTACATTCCTCCGCAAGT 2103
Qy 661 lCysGluAspSerileThrValGlyMetValArgHisGlnGlyLysileMetTyrlValG 681
Db 2104 CTGTGCAGAGTCCATCGTGTGGAACTGTTAAAGCATGAAGGCCAGATCAATATGTCAA 2163
Qy 681 yAspValArgSerValThrGlnLysHisileGlnGlyLleGlyProPheAspLeuVal11 701
Db 2164 TGACGTTCGGAAATACCAAGAAATATTGAAGAGTGGGGCCCGTTCGACTTGGTGAT 2223
Qy 701 eGlyGlySerProCysAsnAspLeuSerileValAsnProAlaArgGlyLleuTyrlG 721
Db 2224 TGGTGAAGCCCATGCAATGATCTCTTAACGTCAATCTGCGCCGCAAGGTTTATATGA 2283
Qy 721 uGlyThrGlyArgLeuPhePheGlyLleuPheTyrlArgLeuLeuHisAspAlaArgProlyG 741
Db 2284 GGGCACAGAAAGGCTCTTCTGAGGTTTACCACCTGCTGAATATATACCGCCGCCAAGGA 2343
Qy 741 uGlyAspAspArgProPhePheThrPhePheGluAsnValValAlaMetGlyValSerAs 761
Db 2344 GGGCGACAAACGTCCTCTTCTGAGTGTGAGAAATGTTGTGGCCATGAAGTGAATGA 2403
Qy 761 pLyAspAspIleSerArgPheLeuGluSerAsnProValMetileAspAlaLysGluVa 781
Db 2404 CAAGAAAGACATCTCAAGATTCCTGGCATGTAAACCCAGTGCATGATGCATGCCATCAAGT 2463
Qy 781 lSerAlaAlaHisArgAlaArgTyrlPheThrPheValAsnLeuProGlyMetAsnArgProle 801
Db 2464 GTCTGCTGCTCACAGGCGCGGTACTTCTGGGTAACTTACCCGGAATGAACAGGCCCGT 2523
Qy 801 uAlaSerThrValAsnAspLysLeuGluGlnGlyCysLeuGluHisGlyValGleAl 821
Db 2524 GATGGCTTCAAGAAATGATAAGCTCGAGCTGCAGGACTGCTCGAGTTCAGTAGGACAGC 2583
Qy 821 alyAspSerileValArgThrileThrArgSerAsnSerileLysGlyLysAs 841
Db 2584 AAGTTAAAGAAAGTGCAGCAATACCAACAGTCCAGATCCATCAGACAGGCAAAHA 2643
Qy 841 pGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTrpCysThrGluMetG 861
Db 2644 CCAAGCTTTTCCCTGTAGTCATGATGGCAAGGACGAGCTTTGTGGTGACATGAGCTGA 2703
Qy 861 uArgValPheGlyPheProValHisTyrlThrAspValSerAsnMetSerArgLeuAlaAr 881
Db 2704 AAGGATCTTCGGCTTCCCTGCTCACTACACGACGCTGTCCAAATGCGCGCGCGCGC 2763
Qy 881 gGlnArgLeuLeuGlyArgSerTrpSerValProValileArgHisLeuPheAlaProle 901
Db 2764 TCAGAAAGTGTGGGAGGTCCTGAGGTGTACCGGTATCCGCTATCAGACACTGTGTTCGCCCTT 2823
Qy 901 uLyGluLysTyrlPheAlaCys 907
Db 2824 GAAGGACTACTTTCCTGT 2842
```

RESULT 11

AF151969

LOCUS

DEFINITION

ACCESSION

4338 bp mRNA linear ROD 08-JUN-2000

Mus musculus DNA cytosine-specific methyltransferase isoform 1

(Dnmt3b) mRNA, complete cds.

AF151969

AF151969.1 GI:8347117

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria;

1 (bases 1 to 4338)

Yin,B., Chen,Y.F., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and

Shen,Y.

Cloning of full-length Dnmt3b cDNA and its alternative splicing

isoforms in mouse embryonic tissue

Unpublished

2 (bases 1 to 4338)

Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and

Shen,Y.

Submitted (17-MAY-1999) Department of Biochemistry and Molecular

Biology, Institution of Basic Medical Sciences, Chinese Academy of

Medical Sciences, Dong Ban Tiao 5, Beijing City 100005,

P.R.China

Location/Qualifiers

1..4338

/organism="Mus musculus"

/strain="KM"

/db\_xref="taxon:10090"

/dev\_stage="8-9 day old embryo"

1..4338

/gene="Dnmt3b"

/feature="Dnmt3b"

/functions="methylates cytosine in DNA"

/note="alternatively spliced"

/codon\_start=1

/product="DNA cytosine-specific methyltransferase isoform

1"

/protein\_id="AAF74515.1"

/db\_xref="GI:8347118"

/translation="MKGDSRLHNEEGAGSYECIIVNGNFSQSSDITKADSPVPVLE

AICTPEVCTPSTRSSRSLSKREVSLLNYTQDMTGDGDDDEVDGSDILMPK

LRTREKTRTRSPAVTRHSNGTSLSLRQASPRITRQGRHHVQYVPEFPATR

SRRRASSTPWSPPASVDMEVTPKSVSTPSVDSQDQGMGTQVDAESIY

GDSYEQDKEFGIGLVMGKIKGFSWMPAMVSVKATKQAMPQMRVOWFGDKF

SEISDKLVALGISOHNLATFNKLSYRKAMVHTLEKARVAGTKTPSSSGESLE

OLKPLMWAHGGFKPTIGLKNPKKOPVNVKSVRRSDSRNLEPRRRNKRRRTTN

DAASESPPPKRLKNTSYGGKDRGDESERMASVETNNKGNLEDRCLSCGKNPVS

PHLPFEGGLQCRCRDFLEFLFYMDGYSYCTVCCGREGILLDSCTSCRCFVPS

LEVLVGATAEADAKLOEPMSCVCLPQCHGLPARRKDMNMLQDFPTDPLSEFEP

PKLYPAIAAKRPPRIKVLSELDGATVYLKELGKIKVEXIASEVCAESIAGTVKH

EQQLKYNDVRKITKNIIEWGFDFLVIGSPCNLSNVNPAKGLYEGTGLTFEYI

HLNLTIRPKGDNRPFFWPNVNAVKVNDKQISRLACNPVMDAIKVSAAHRAHY

FWGLPMNRPVNAKNDKLQDCLCFSTAKLKKVQVITITKSNIRQGNKQLFPVY

MNGKDDVLMCTELERIFGPFHYHTDVSNMGRGAKQKLLGHSVSVPIRHLFAPLKDYF

ACE"

BASE COUNT 1073 a 1124 c 1173 g 968 t

ORIGIN

Alignment Scores:

Pred. No.: 2,71e-90 Length: 4338

Score: 2206.00 Matches: 455

Percent Similarity: 59.78% Conservative: 135

Best Local Similarity: 46.10% Mismatches: 268

Query Match: 44.94% Indels: 129

DB: 10 Gaps: 15

US-09-720-086-5 (1-908) x AF151969 (1-4338)

Qy 4 SerGlyProGlyAspThrSerSerSer-----SerLeuGluArgGluAsp 18

Db 165 AGCGGCCCAAGTAAACGATGACGACGATCGCGCGCGGAGATTCCGGAACCCGACATCC 224

Qy 19 AspArgLysGluGlyGluGlnGlnGluAsnArgGly-LysGluGluArgGlnGluPr 38

Db 225 GCSCGCCCGCGCGGACGACCCCGCGCGGATCGCGCGCGCGGCTACAGCCAGCTCA 284

QY 38 oSerAlaThrAlaArgIuValIglYarProGluYarGlyValGlyHisPro--ProVa 57  
DB 285 CGACAGCGCCGCTGAGAGCTTGCCAGACCTTGAAACCTCAGATATATCTTCCAGA 344  
QY 57 IgluSerSerAspThrProIyAspProAlaValThrThrIleSerGlnProMetAlaIgl 77  
DB 345 CGGGGATCTCCCTCCCTCCATCATGCTTGGAACCAATCCAGGGCTTTCTTCA 404  
QY 77 rAspSerGlyProSerAsp----- 83  
DB 405 GGAACAATGAAGGAGACAGACACATGTGATGAAGAAGAGGTGCCCGGGATGA 464  
QY 84 -----LeuLeuProAsnGlyAspLeuGluYarGlySerGlnProGluGluGlu 101  
DB 465 GGAGTCATTATGTTAATGGGAATTCAGTGAACAGTCTCTCAGACACGAAAGATGCTCC 524  
QY 101 ySerProAlaAlaGlyGlnIuValIglYalProAlaGluGluGlyThrGluThrPr 121  
DB 525 CTCAACC-----CC 533  
QY 121 oProGluAlaSerArgAlaValIgluAsnGlyCysCysValThrIleGluGluYarGlyAl 141  
DB 534 AGCTTGAGAGGCATCTGCACAGCCAGCTCTGCACACAGACAGAGGCGCCAGCTC 533  
QY 141 aSerAla----- 143  
DB 594 AAGCTCCCGCTGTCTAAGAGGAGGCTCCAGCTTCTGAATTACAGCAGACATGAC 653  
QY 144 -GlyGluGly----- 146  
DB 654 AGGAGATGAGACAGAGATGATGAAGTGAATGGATGCTGTGATATTCTTAATGCC 713  
QY 147 -----LysGluGluIuValGlnThrAsnIleGluSerMetIleGluGlySerAr 163  
DB 714 AAGCTCACCCCTGAGACACCAAGACACAGACCGCTCT-----GAAACCCCGGC 764  
QY 163 gGlyArgLeuArgGlyGlyLeuGlyYTrpGluSerSerLeuArgGluArgProMetProAr 183  
DB 765 TGTCCGAACCCGACATAGCAATGGAGACCTCCAGCTTGAGAGGCAAAAGACCTCCCCAG 824  
QY 183 gLeuThrPheGlnAlaGlyAspProTyTrIleSerIyAspGlyArgAspGluTrpLe 203  
DB 825 AATCACCCGAGGCTGGCAGGCGCCACCATGTG-----CAGAGATGCC 869  
QY 203 uAlaAGTrpIyAspGluAlaGluIuValIleValIleAlaValMetAsnAlaVa 223  
DB 870 TGTGGAGTTCCCGCTACCAAGCTTCGAGAGCTCGAGCATCGTCTTCAGCAAGCAGCC 929  
QY 223 IgluIuAsnGlnAlaSerGlyIuSerGlnIuValIgluAlaIleSerProProAlaVa 243  
DB 930 ATGTCATCCCTCCGACGCTGCATCTC-----ATGGAAGAAGTGAACACTMAAGCGT 983  
QY 243 IgluGlnProThrAspProAlaSerProThrValAlaThrProGluProValGlyI 263  
DB 984 CAGTACCCCA-----TCAGTTGACTTGAGCCAGAGTGAATGAGAGGCTAT 1031  
QY 263 yAspAlaGlyAspIyAsnAlaThrIyAlaAlaAspAspGluProGluIyTrpGluAspG 283  
DB 1032 GGAATACCAACAGAGTGAATGACAGACATATATGAGACAGACAGAGATACAGATGA 1091  
QY 283 yArgGlyPheGlyIleGlyIuLeuValITrpGlyIyLeuArgGlyPheSerTrpTrp 303  
DB 1092 TAAAGAGTTTGAATAGTGAACCTCGTGTGGGAAGAATGAAGGCTTCTCGTGGTGGC 1151  
QY 303 oGlyArgIleValIleSerTrpTrpMetThrGlyArgSerArgAlaAlaGluGlyThrArgTr 323  
DB 1152 TGCATGATGTGTCTCTGGAAGCACTCCAGCGACAGGCAATGCCGGAATGCCCTG 1211  
QY 323 pValMetTrpPheGlyAspGlyIyAspPheSerValIleCysValIgluIyLeuMetProLe 343  
DB 1212 GGTACAGTGTGTGGTGAAGAGTTTCTGAGATCTCTGCTGACAACTGGTGGCTCT 1271

QY 343 uSerSerPheCysSerAlaPheHisGlnAlaThrTyAsnIyGlnProMetTyArgIy 363  
DB 1272 GGGGCTGTTCAGCCAGACACTTAACTGTGCTACCTTCAATAAGCTGTTCTTATAGAA 1331  
QY 363 sAlaIleTyArgIuValIleGluValAlaSerSerArgAlaGlyIyLeuPheProAlaCy 383  
DB 1332 GGCATGTACACACTCTGAGAAACCGAGGTTCCAGCTGGCAACCTTC----- 1383  
QY 383 sHisAspSerAspGluSerAspSerGlyIyAlaValIgluAlaGlnAsnIyGlnMetI 403  
DB 1384 -----TCCAGAGCTCTGAGAGTCACTCGAGAGCAACAGCTGAAGCCATGCT 1430  
QY 403 eGluTrpAlaLeuGlyIyPheGlnProSerGlyProIyGlyLeuGluPro----- 420  
DB 1431 GGAGTGGGCCCAACGGTGTCAAGCTTCTGGATCGAGGGCTTCAAAACCAACAGAA 1490  
QY 421 -----ProG 422  
DB 1491 GCACACAGTGTTAATAGTCGAAGGTGCTGCTTCAGACAGTGAAGACTTGAACCCAG 1550  
QY 422 uGluGluIyAsnProTyTrIleGluValTyThrAspMetTrpValIgluProGluAlaI 442  
DB 1551 GAGACCGAGAACAAAGTCGAAGAGCGCAACCAAT-----GACTCTGC 1595  
QY 442 aAlaTyAlaProProProProAlaIyIyAspProArgIySerThrThrGluIyProIy 462  
DB 1596 TGCTTGTGATGTCCTCCCAACCAAGCGCTCAAGACAAATAGATATGGCGGAAGAGCG 1655  
QY 462 sValIySGluIleIleAspGluArgThrArgGluValIyGluValAlaArgGluIy 482  
DB 1656 AGGGAG-----GATGAGGAGCGCGAGACGAGATGCTTGAAGTCAACAA 1706  
QY 482 sCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThrLeuGluH 502  
DB 1707 CAAGGCAATCTGGAAGACCGCTGTTCTGTGAAAGAAAGAAACCTGTGCTTCCA 1766  
QY 502 sProLeuPheIleGlyIyMetCysGlnAsnCysIyAsnCysPheLeuGluCysAlaTy 522  
DB 1767 CCCCTCTTGAAGGTGGGCTGTGACAGTGGCCGAGTCCGCTTCTAGAGCTTTCTA 1826  
QY 522 rGlnTyAspAspAspGlyTyTrpGlnSerTyCysThrIleCysCysGlyIyArgGluVa 542  
DB 1827 CATGTATGATGAGAGGATATCATGCTACGACCGGTGTGTGAGGGCGGTAACT 1886  
QY 542 IleuMetCysGlyIyAsnAsnAsnCysArgCysPheCysValIgluCysValAspLeu 562  
DB 1887 GCTGTGTGAGTACACAAAGCTGTGCAGATGCTTCTGTGTGAGTGTGTGAGGTGCT 1946  
QY 562 uValIgluProGlyAlaAlaGlnAlaAlaIleIySGluAspProTrpAsnCysTyMetCy 582  
DB 1947 GGTGGCGCAGGACAGCTGAGGATGCCAAAGTCGAGAAACCTGAGGCTGTATATGTG 2006  
QY 582 sGlyHisIySGlyThrTyGlyIleuLeuArgArgGluAspTrpProSerArgLeuG 602  
DB 2007 CCTCCCTCAGCGCTGCATAGGGGTCTCCGACGCAAGAAAGATTGGAACATGCGCTGCA 2066  
QY 602 pMetPhePheAlaAsnAsnHisAsp-----GlnGluPheAspProIyValIyTrpPr 621  
DB 2067 AGACTTCTTCACTACTGATCTCTAAGCTTGAAGATTTGAGCCACCCAGATGATCCAGC 2126  
QY 621 oValProAlaGluIyAspIyAspProIleArgValIleuSerLeuPheAspGlyIleAlaTh 641  
DB 2127 AATTCCTGAGCCAAAGAGAGGCCATTGAAGTCCGTCTGTGTTGATGAATTCGAAC 2186  
QY 641 rGlyLeuLeuValIleuIyAspLeuGlyIleGlnValAspArgTyTrIleAlaSerGluVa 661  
DB 2187 GGGGTACTTGCTGCTCAAGAGTGGGTATTAAGTGAAGAAAGTATGATGCCCTCCAAAG 2246  
QY 661 IySGluAspSerIleThrValIgluMetValIArgHisGlnGlyIyIleMetTyTrpAl 681  
DB 2247 CTGTGCAAGTCCATGCTGTGTGAACCTGTGAAGCTGAAGGCGCATGAATATGTCAA 2306  
QY 681 yAspValArgSerValThrGlnIyHisIleGlnIuTrpGlyProPheAspLeuValI 701

DB 2307 TGACGTCGCGAAATACCAAGAAATATTGAAGAGTGGGCGCGTTCGACTTGGTAT 2366  
QY 701 eGlyGlySerProCysAsnAspLeuSerIleValAsnProAlaArgIleValGlyLeuTyG1 721  
DB 2367 TGGTGGAGAGCCATCGAATGATCTCTTAACGTGTAATCTTCCGCGCAAGGTTTATATGA 2426  
QY 721 uGlyThrGlyArgLeuPhePheGluPheTyArgLeuLeuLeuHisAspAlaArgProIleG1 741  
DB 2427 GGGCACAGGAAGGCTCTCTTCGAGTTTACCACCTGCTGAATATATACCGCGCCCAAGGA 2486  
QY 741 uGlyAspAspArgProPhePheTrpLeuPheGluAsnValValAlaMetGlyValSerAs 761  
DB 2487 GGGCGCAACACCGTCCATCTCTTCTGGATGTTGAGAAATGTTGTGGCCATGAAGTGAATGA 2546  
QY 761 pIyAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaIyGluVa 781  
DB 2547 CAAGAAAGACATCTCAAGATCTCTGGCATGTAAACCCAGTGTATGATGATGCCATCAAGGT 2606  
QY 781 lserAlaAlaHisArgAlaArgTyPheTrpGlyAsnLeuProGlyMetAsnArgProIle 801  
DB 2607 GTCTGCTGCACAGGCGCGGTACTTCTGGGGTAACTACCGGAATGAACAGGCGCGT 2666  
QY 801 uAlaSerThrValAsnAspIleLeuGluLeuGluCysLeuGluHisGlyValIleAl 821  
DB 2667 GATGCTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2726  
QY 821 alyPheSerIyValArgThrIleThrArgSerAsnSerIleIyGluGlyAs 841  
DB 2727 AAAGTTAAAGAAAGTGCAGACAATAACCAACAGTCGACATCCATCAGACAGGCGCAAAA 2786  
QY 841 pGlnHisPheProValPheMetAsnGluIyGluAspIleLeuTrpCysThrGluMetG1 861  
DB 2787 CCAGCTTTTCCCTGTGTATGATCATGAATGGCAAGACAGCGTTTGTGGTCACTGAGCTGA 2846  
QY 861 uArgValPheGlyPheProValHisTyThrAspValSerAsnMetSerArgLeuAlaAr 881  
DB 2847 AAGGATCTTCGGCTTCCTGCTCACTACACGAGCGTGTCCAACTGGCGCGCGCGCG 2906  
QY 881 gGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAlaProIle 901  
DB 2907 TCAGAAAGTCTGGCGAGGCTCTGGAGTGTACCGGTGATCATCAGACACCTGTTTGGCCCTT 2966  
QY 901 uIyGluTyPheAlaCys 907  
DB 2967 GAAGGACTACTTTCCTGT 2985  
RESULT 12  
AF151974  
LOCUS  
DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 6  
ACCESSION AF151974  
VERSION AF151974.1  
KEYWORDS GI:8347130  
SOURCE Mus musculus.  
ORGANISM Mus musculus  
REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 4163)  
Yin, B., Chen, Y.T., Zhu, M., Luo, Y.J., Xu, S.C., Wu, G.Y. and  
Shen, Y.  
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing  
isoforms in mouse embryonic tissue  
JOURNAL Unpublished  
REFERENCE  
AUTHORS 2 (bases 1 to 4163)  
Yin, B., Chen, Y.T., Zhu, M., Luo, Y.J., Xu, S.C., Wu, G.Y. and  
Shen, Y.  
TITLE Direct Submision  
JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular  
Biology, Institution of Basic Medical Sciences, Chinese Academy of  
Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,  
P.R.China

FEATURES  
source  
Location/Qualifiers  
1. 4163  
/organism="Mus musculus"  
/strain="KM"  
/db\_xref="taxon:10090"  
/dev\_stage="8-9 day old embryo"  
1. 4163  
/gene="Dnmt3b"  
297..2816  
/function="methylates cytosine in DNA"  
/note="alternatively spliced"  
/codon\_start=1  
/product="DNA cytosine-specific methyltransferase isoform 6"  
/protein\_id="AAF74520.1"  
/db\_xref="GI:8347131"  
/translation="MKGDSRLHNEEGASGYEECIIVNGNFSQSSDTTKDAFSPV  
AICTEPVCTPETERSSRLSKREVSLNLTQMDGDRDDDEVDGSDILMPK  
LRTREKDTTRSPAPVTRHSNGTSSLERQASPRITRQGRHGVQVYPVDFPATR  
SRRRASSTASFPWSPASVDMEEVTPKSVTSPDLSDQDQGGMDTQVDFASIT  
GDSTEYQDDKFEIGLGVNMGKIKGFPMPAMVSMKATKQAMPQMRVWFGGKIF  
SEISADLVALGLFSOHLVATFNKLVSRYKAMVHTLEKARVAGTTFSSPGESELD  
OLKMLEAHGKFPKPTGIEGLPKKKOPENKRRRTNDSAASEPPKRLKTNYSYGG  
KRGEDSESRNMASEVTNNKGNLDRCLSCGKNPVSPHPLFEGGLCOSCRDRLEL  
PYMDENGYOCTVCCGRELLOFTTDPDLSEPEPKLYPAIAKRRPTLVLSL  
CYMCLPORCHGLRRKDNWMLQFTTDPDLSEPEPKLYPAIAKRRPTLVLSL  
FDGATATYLVKELGIVKVEYIASVCAESIAVGTVEGQIKVNDVSKIKKNIIE  
WGFDDLVIGSPDNLSNVPARKGLYEGLTFEYFVHLLNVTREKGDNRPFWMP  
ENVYAMVNDKDI SRFLACNPVMDAIKYSNAHRAKRYFNGNLFGMRPFWASKDKL  
ELQCLFESRTAKLKKQLTITKSNIRQGGKLPFPVVMGKDDVLWTELEKIFGFP  
AHYTDVSNMGRGAKKLLGRSWSVPVIRHLFAPLKDIFYACE"  
BASE COUNT 1032 a 1074 c 1128 g 929 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3.37e-90 Length: 4163  
Score: 2203.50 Matches: 448  
Percent Similarity: 61.91% Conservative: 134  
Best Local Similarity: 47.66% Mismatches: 236  
Query Match: 44.89% Indels: 123  
DB: 10 Gaps: 15  
US-09-720-086-5 (1-908) x AF151974 (1-4163)  
QY 44 LysValGlyArgProGlyValGlyArgLys----- 53  
DB 130 CGACAGCGGCTCCAGTCGCGCGCGGGTTAAGCGGCCCAAGTAAAGTAGCGAG 189  
QY 54 -----HisProProValGluSerSerAspThrProLysAspProAlaValThrThyls 71  
DB 190 CGATCGCGCGGAGATTCGGCAACCGCACCTCCGCGCGCGCGCGCGCGCGCGCG 249  
QY 72 -----SerGlnProMetAlaGlnAspSerGlyProSerAsp 83  
DB 250 GGCGCGATCGCGCGCGCGCTACAGCCAGCCTCA-CGACAGGAACAACAAAGGAGAC 308  
QY 84 -----LeuLeuProAsn 87  
DB 309 AGCAGACATCTGAATGAAGAAGAGGGTGCAGCGGGGTATGAGGAGTGCTATTCGTTAAT 368  
QY 88 GlyAspLeuGluIyArgSerGluProGlnProGluGluGlySerProAlaAlaGlyGln 107  
DB 369 GGGAACTTCAGTACAGGAGTCTCTCAGACACAGAGGAGTCTCCTCACCC----- 416  
QY 108 LysGlyAlaProAlaGluGlyGluGlyThrProProGluAlaSerArgAla 127  
DB 417 -----CCAGTCTTGGAGGCAATCTGC 437  
QY 128 ValGluAsnGlyCysValThrIyGluGlyArgGlyAlaSerAla----- 143  
DB 438 ACAGAGCCAGTCTGCACACACAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 497

QY 144 -----GlyGluGly----- 146  
 Db 498 AGGAGAGTCTCCAGCTTCTGAAATTAACGACGACATGACAGAGATGAGACAGAGAT 557  
 QY 147 -----LysGluGln 149  
 Db 558 GATGAAGTATGATGGAATGGCTCTGATATTTCTAATGCGAAAGCTCACCCGTGAGACC 617  
 QY 150 LysGlnThrAsnIleGlnSerMetLysMetGluGlySerArgGlyValArgLeuArgGlyGly 169  
 Db 618 AAGGACACACGACGCGCTCT-----GAAAGCCCGGCTGTCGGAACCCGACATAGC 668  
 QY 170 LeuGlyTyrGlnSerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGly 189  
 Db 669 AATGGACCTCCAGCTTGAGAGGACAAAGGCTCCCGCGAATCAACCCAGGTCCGACG 728  
 QY 190 AspProTyrTyrIleSerLysArgLysArgAspGluTyrPheLeuAlaArgTyrLysArgGlu 209  
 Db 729 GGGCGCCACCATGTG-----CAGAGTACCTCTGTGAGTTTCCGCGCTACC 773  
 QY 210 AlaGluLysLeuAlaLysValIleAlaValMetAsnAlaValGluGlnSerGlnAlaSer 229  
 Db 774 AGGTCTCGAGACGTGAGAGATGTCTTCAAGACACGCGCATGTATCCCTGCGACG 833  
 QY 230 GlyGlnSerGlnLysValGluGlnAlaSerProProAlaValGlnGlnProThrAspPro 249  
 Db 834 GTGACCTTC-----ATGGAGAAGTGAACCTTAAGAGCTCAGTACCCCA----- 878  
 QY 250 AlaSerProThrValAlaThrThrProGluProValGlyGlyAspAlaGlyAspLysAsn 269  
 Db 879 ---TCAGTTGACTTGAGCCAGATGAGATGACGAGGGATGATGCCACACAGGTGAT 935  
 QY 270 AlaThrLysAlaAlaAspAspGluProGluTyrGluAspGlyValArgGlyPheGlyIleGly 289  
 Db 936 GCAGAGACGATATATGAGACGACACAGATATCAGAGATGATTAAGAGTTGGAATAGCT 995  
 QY 290 GlnLeuValTyrGlyLysLeuArgGlyPheSerTyrTyrProGlyValArgIleValSerTyr 309  
 Db 996 GACCTGTGTGGGGAAGATCAAGGCTTCTCTGTGGCTCCATGTGGTGTCTCTGG 1055  
 QY 310 TrpMetThrGlyArgSerArgAlaAlaGluGlyThrArgTyrValMetTyrPheGlyAsp 329  
 Db 1056 AAGGACCTCCCAAGGACGACGCGCATGCCGGAATGCGTGGTAAAGTGTGTGAT 1115  
 QY 330 GlyLysPheSerValValCysValGluLysLeuMetProLeuSerSerPheCysSerAla 349  
 Db 1116 GGCAATTTTTCAGATCTCTGCTGACAACTGTGCTTGGGCTGTTCACCCAGCAC 1175  
 QY 350 PheHisGlnAlaThrTyrAsnLysGlnProMetTyrArgLysAlaIleTyrGluValLeu 369  
 Db 1176 TTTAATCTGGCTACTTCATATAGCTGTCTTATTAGGAAGGCCATGTACACACTCTG 1235  
 QY 370 GlnValAlaSerSerArgAlaGlyLysLeuPheProAlaCysHisAspSerAspGlnSer 389  
 Db 1236 GAGAAAGCCAGGGTTGAGCTGCAAGACTTC-----TCCAGC 1274  
 QY 390 AspSerGlyLysAlaValGluValGlnAsnLysGlnMetIleGlnTyrAlaLeuGlyGly 409  
 Db 1275 AGTCCTGGAGAGTCACTGAGGACAGCTGAAACCCATGCTGAGTGGGCGCCAGGGGCG 1334  
 QY 410 PheGlnProSerGlyProLysGlyLeuGlnProProGluGlnGlu---LysAsnProTyr 428  
 Db 1335 TTCACCTTACTGGATCGAGGGCTCAAAACCAAGAGACCCGACGAAACAAAG 1394  
 QY 429 LysGluValTyrThrAspMetTyrValGluProGluAlaAlaAlaTyrAlaProPro 448  
 Db 1395 CGAAGACGACCAACAT-----GACTCTGCTCTCTGAGTCCCCCA 1439  
 QY 449 ProAlaLysLysProArgLysSerThrThrGluLysProLysValLysGluIleLeuAsp 468  
 Db 1440 CCCAAGCGCTCAAGAACAAATAGCTATGGCGGAGAGACCGAGGGAG-----GAT 1490  
 QY 469 GlnArgThrArgGlnArgLeuValTyrGluValArgGlnLysCysArgAsnIleGluAsp 488

Db 1491 GAGAGAGCGCGAAGACGATGGCTTCTGAAGTACCAACAAAGGGCAATCTGGAGAGC 1550  
 QY 489 IleCysIleSerCysGlySerLeuAsnValThrLeuGlnHisProLeuPheIleGlyGly 508  
 Db 1551 CGCTGTTTCTCTGTGGAAGAAAGAACCTGTGTCTCTTCCACCCCTCTTTAGAGGTGG 1610  
 QY 509 MetCysGlnAsnCysLysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAspArgGly 528  
 Db 1611 CTCTGTCAAGATGTGCGGATGCGCTTCTTAAGAGCTTCTTCAATGATATGATGAGCGGC 1670  
 QY 529 TyrGlnSerTyrCysThrIleCysCysGlyValArgGluValLeuMetCysGlyAsnAsn 548  
 Db 1671 TATCACTCTTACTGACCCGTGTGCTGTGAGGCGCGGAACCTGCTGTGTCAGTACACA 1730  
 QY 549 AsnCysCysArgCysPheCysValGluCysValAspLeuLeuValGlyProGluAlaAla 568  
 Db 1731 AGCTGCTGAGATGCTTCTGTGTGAGTGTCTGTGAGGTCTGTGTGGCCGACGACAGCT 1790  
 QY 569 GlnAlaAlaIleLysGluAspProTyrAsnCysTyrMetCysGlyHisLysGlyThrTyr 588  
 Db 1791 GAGATGCCAAGCTGAGAGAACCTGAGACTGTATGTGCTCCCTCAGCGCTGCCAT 1850  
 QY 589 GlyLeuLeuArgArgGluAspTyrProSerArgLeuGlnMetPhePheAlaAsnAsn 608  
 Db 1851 GGGGTCTCCGACGACGAAAGATTGGAACATGCGCTCAAGACTTCTTCACTACTGAT 1910  
 QY 609 HisAsp---GlnGluPheAspProProLysValTyrProProValProAlaGluLysArg 627  
 Db 1911 CCTGACTGGAAGATTTAGCCACCCCAAGTTGATCCCGCAATTCCTGACGCAAAAG 1970  
 QY 628 LysProIleArgValLeuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeuLys 647  
 Db 1971 AGGCCATTAGAGTCTGTCTCTGTGATGGAATTGCAACGGGACTTGTGCTCAAG 2030  
 QY 648 AspLeuGlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIleThr 667  
 Db 2031 GAGTTGGGTATTAAAGTGAAAGATACATGCTCCGAGTCTGTGCAAGATCCATCGCT 2090  
 QY 668 ValGlyMetValArgHisGlnGlyLysIleMetTyrValGlyAspValArgSerValThr 687  
 Db 2091 GTGGGAACCTTTAAGCATGAAGCCAGATCAATATGATATGAGGGACAGAAAGGCTTTC 2150  
 QY 688 GlnLysHisIleGlnGluTyrGlyProPheAspLeuValIleGlyLysSerProCysAsn 707  
 Db 2151 AAGMAAAATATTGAAGAGTGGGGCCGTTGCACTTGATTTGGTGAAGCCATGCAAT 2210  
 QY 708 AspLeuSerIleValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhe 727  
 Db 2211 GATCTCTTAACGTCAATCTGCGCGAAAGTTATATGAGGGACAGAAAGGCTCTTC 2270  
 QY 728 PheGluPheTyrArgLeuLeuHisAspAlaArgProLysGluGlyAspAspArgProPhe 747  
 Db 2271 TTCGAGTTTACCACTGTCTGTAATTATACCCGCCCAAGAGGGGCAACACCTGCATTC 2330  
 QY 748 PheTyrPhePheGlnAsnValValAlaMetGlyValSerAspLysArgAspIleSerArg 767  
 Db 2331 TTCTGATTTGAGAAATGTGTGGCATGAAGATGAATGACAAAGAAACATCTCAAGA 2390  
 QY 768 PheLeuGlnSerAspProValMetIleAspAlaLysGluValSerAlaAlaHisArgAla 787  
 Db 2391 TTCCTGGCATGAACCAAGTATGATGATGCAATCAAGTGTCTGCTCACAGGGCC 2450  
 QY 788 ArgTyrPheTyrGlyAsnLeuProGlyMetAsnArgProLeuAlaSerThrValAsnAsp 807  
 Db 2451 CGTACTTCTGGGTAACCTTACCCGAATGAACAGGCCGTGATGCTTCAAAAGATATGT 2510  
 QY 808 LysLeuGlnLeuGlnCysArgLeuGlnHisGlyArgGlyAlaLysPheSerLysValArg 827  
 Db 2511 AAGCTGAGCTGAGAGCTGCTGAGATTCAATGAGACGCAAAAGTTAAAGAAAGTGCAG 2570  
 QY 828 ThrIleThrThrArgSerAsnSerIleLysGlnGlyLysAspGlnHisPheProValPhe 847

Db 2571 ACAATAACCAACCAAGTCGAACCTCCATCAGACAGGCGAAAAACCAAGCTTTTCCTCTGATC 2630  
QY 848 MetAenGluLysGluAspIleLeuTrpCysThrGluMetGluArgValPheGlyPhePro 867  
Db 2631 ATGAATGCAAGGACGACGCTTTTGTGGTCCACTGAGCTCGAAAGGATCTTCGGCTTCCT 2690  
QY 868 ValHisTrpThrAspValSerAsnMetSerArgLeuAlaArgGlnArgLeuLeuGlyArg 887  
Db 2691 GCTCCTACACGACGAGCTGTCCAAACATGGCGCGCGCCCTGAGAGCTGTCTGGCGCAGG 2750  
QY 888 SerTrpSerValProValIleRghIleLeuPheAlaProLeuLysGluTrpPheAlaCys 907  
Db 2751 TCTTGGAGTGTACCGGTCTATCAGACACCTGTTTCCCTTCCTGAGGACTACTTTCCTGT 2810  
RESULT 13  
AC112586/c  
LOCUS AC112586 176697 bp DNA linear HTG 13-JUL-2002  
DEFINITION Rattus norvegicus clone CH230-28122, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 65 unordered pieces.  
ACCESSION AC112586  
VERSION AC112586.3 GI:21738467  
KEYWORDS HTG; HTGS PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 176697)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,  
Alabrooke, S.L., Amaral, J., Amaratunga, H.C., Are, J.R., Ayele, M., Banks, T.,  
Barbora, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Bowie, S., Brieve, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Delaney, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Deltany, K.R., Delgado, O., Denna, L.L., Ding, Y., Dinh, H.H.,  
Douthett, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Eamthair, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Franz, P.,  
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, K.,  
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,  
Homes, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
Karlssohn, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Lousaged, H.,  
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
Magesy, E., Mahoney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Mosier, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okuwon, G.,  
Orgunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,  
Rivers, M., Rojas, A., Rodubokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
Sodergren, E., Sonake, I., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalobos, B., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wlecczyk, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.  
Direct Submission  
TITLE Unpublished  
REFERENCE 2 (bases 1 to 176697)  
AUTHORS Worley, K.C.  
TITLE Direct Submission

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Submitted (22-FEB-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 176697)  
Worley, K.C.  
Direct Submission  
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 12, 2002 this sequence version replaced gi:20303171.

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
Project Information  
Center project name: GRAT  
Center clone name: CH230-28122  
Summary Statistics  
Sequencing vector: Plasmid  
Chemistry: Dye-terminator Big Dye: 100% of reads  
Assembly program: Phrap: version 0.990329  
Consensus quality: 121959 bases at least Q40  
Consensus quality: 126371 bases at least Q30  
Consensus quality: 130428 bases at least Q20

NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
NOTE: This is a 'working draft' sequence. It currently  
consists of 65 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 1415: contig of 1415 bp in length  
1416: gap of unknown length  
1516: contig of 1450 bp in length  
2966: gap of unknown length  
3065: gap of unknown length  
4226: contig of 1163 bp in length  
4229: gap of unknown length  
5385: contig of 1057 bp in length  
5386: gap of unknown length  
5486: contig of 1185 bp in length  
6671: gap of unknown length  
7813: contig of 1043 bp in length  
7913: gap of unknown length  
7914: contig of 1760 bp in length  
9674: gap of unknown length  
9774: contig of 1697 bp in length  
11470: gap of unknown length  
11471: gap of unknown length  
11570: contig of 1592 bp in length  
13162: gap of unknown length  
13163: gap of unknown length  
13263: contig of 1242 bp in length  
14504: gap of unknown length  
14505: gap of unknown length  
14605: contig of 1370 bp in length  
15975: gap of unknown length  
16075: contig of 1322 bp in length  
17397: gap of unknown length  
17496: contig of 1827 bp in length  
19323: gap of unknown length  
19324: gap of unknown length  
21043: contig of 1618 bp in length  
21042: gap of unknown length  
22413: contig of 1278 bp in length  
22420: gap of unknown length  
23420: contig of 1423 bp in length  
23942: gap of unknown length  
24042: gap of unknown length  
25333: contig of 1291 bp in length  
25334: gap of unknown length  
25433: contig of 1097 bp in length  
26530: gap of unknown length  
26531: contig of 1640 bp in length



\* 28371 28370: gap of unknown length  
\* 28371 28370: contig of 1600 bp in length  
\* 29971 30070: gap of unknown length  
\* 30071 32265: contig of 2195 bp in length  
\* 32265 32365: gap of unknown length  
\* 32365 34179: contig of 1814 bp in length  
\* 34180 34279: gap of unknown length  
\* 34280 36392: contig of 2113 bp in length  
\* 36393 36492: gap of unknown length  
\* 36493 38552: contig of 2060 bp in length  
\* 38553 38652: gap of unknown length  
\* 38653 40932: contig of 2280 bp in length  
\* 40933 41032: gap of unknown length  
\* 41033 43021: contig of 1989 bp in length  
\* 43022 43121: gap of unknown length  
\* 43122 44977: contig of 1856 bp in length  
\* 44978 45077: gap of unknown length  
\* 45078 47358: contig of 2281 bp in length  
\* 47359 47558: gap of unknown length  
\* 47459 49554: contig of 2096 bp in length  
\* 49555 49654: gap of unknown length  
\* 51876 51876: contig of 2222 bp in length  
\* 51877 51976: gap of unknown length  
\* 51977 53794: contig of 1818 bp in length  
\* 53795 53894: gap of unknown length  
\* 53895 55634: contig of 1740 bp in length  
\* 55635 55734: gap of unknown length  
\* 55735 57445: contig of 1711 bp in length  
\* 57446 57545: gap of unknown length  
\* 57546 60210: contig of 2665 bp in length  
\* 60211 60310: gap of unknown length  
\* 60311 62174: contig of 1864 bp in length  
\* 62175 62274: gap of unknown length  
\* 62275 64285: contig of 2011 bp in length  
\* 64286 64385: gap of unknown length  
\* 64386 66764: contig of 2379 bp in length  
\* 66765 66864: gap of unknown length  
\* 66865 69398: contig of 2534 bp in length  
\* 69399 69498: gap of unknown length  
\* 69499 71451: contig of 1953 bp in length  
\* 71452 71551: gap of unknown length  
\* 71552 74405: contig of 2854 bp in length  
\* 74406 74505: gap of unknown length  
\* 74506 76586: contig of 2081 bp in length  
\* 76587 80397: gap of unknown length  
\* 80398 80497: gap of unknown length  
\* 80498 82299: contig of 1802 bp in length  
\* 82300 82399: gap of unknown length  
\* 82400 84796: contig of 2397 bp in length  
\* 84797 84896: gap of unknown length  
\* 84897 86981: contig of 2085 bp in length  
\* 86982 87082: gap of unknown length  
\* 87082 89794: contig of 2713 bp in length  
\* 89795 89894: gap of unknown length  
\* 89895 91696: contig of 1802 bp in length  
\* 91697 91796: gap of unknown length  
\* 91797 94458: contig of 2662 bp in length  
\* 94459 94558: gap of unknown length  
\* 94559 98352: contig of 3794 bp in length  
\* 98353 98452: gap of unknown length  
\* 98453 101134: contig of 2682 bp in length  
\* 101135 101234: gap of unknown length  
\* 101235 104588: contig of 3354 bp in length  
\* 104589 104688: gap of unknown length  
\* 104689 107761: contig of 3073 bp in length  
\* 107762 107861: gap of unknown length  
\* 107862 111340: contig of 3479 bp in length  
\* 111341 111440: gap of unknown length  
\* 111441 115374: contig of 3934 bp in length

Alignment Scores:

2.04e-88

Length:

176697

Score: 2200.50 Matches: 415  
Percent Similarity: 89.86% Conservative: 19  
Best Local Similarity: 85.92% Mismatches: 39  
Query Match: 44.83% Indels: 11  
DB: 2 Gaps: 2

US-09-720-086-5 (1-908) x AC112586 (1-176697)

QY 391 SerGlyLysAlaValGluValGlnAenLysGlnMetIleGluTrpAlaLeuGlyGlyPhe 410  
DB 23941 TCTCGAACGGGTGTGAGGTGATTAACCTA-TTGCCATTGATGGGTCTTGGCGGCTTC 23883  
QY 411 GlnProSerGlyProLysGlyLeuGluProProGluGluGlyLysAsnProTrpLysGlu 430  
DB 23882 CAGCCTTGTCGCCAAGGCGCTGTAGCCACCATTAAGAGAAACAATCTTACAAAGAA 23823  
QY 431 ValTrpThrAspMetTrpValGluProGluAlaAlaIleTyrAlaProProProAla 450  
DB 23822 GTTTACACCGAATGTGGGTGAGCCTGAGGAGCTGATTATTCGCCGCCACAGCC 23763  
QY 451 LysLysProArgLysSerThrTrpLysLys-ProLysValLysGluIleIleAspGluAr 470  
DB 23762 AACAAACCCAGAAAGAGC-----ACAACTTAAGGTCAAGAGATATTGATTAACA 23712  
QY 470 GThrArgGluArgLeuValTyrGluValArgLysLysCysArgAsnIleGluAspIleCy 490  
DB 23711 CCGAAGATACGACCTATGATGAGGTGCCAAAGTCCAAAGCATGAGACATTTTG 23652  
QY 490 GlleSerCysGlySerLeuAsnValThrLeuGluHisProLeuPheIleGlyGlyMetCy 510  
DB 23651 TATTCATGTGGAGGCTCAAGTCAACCTGAGAGCAACCACTTCATTGTGGGAATGTG 23592  
QY 510 GlnAenCysLysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAspAspGlyTyrGly 530  
DB 23591 CCACAACTTAAGAACTGCTTCTTGAGTGTGCTTACAAATACGATGACATGTGTACA 23532  
QY 530 nSerTrpCysThrIleCysCysGlyGlyArgGluValLeuMetCysGlyAsnAsnAsnCys 550  
DB 23531 GTCCCTTGATCATCTGCTGTGGGGGCGGAAGTCTCAATGTGTGGGAACAACACCG 23472  
QY 550 GcysArgCysPheCysValGluCysValAspLeuLeuValGlyProGlyAlaAlaGlnAl 570  
DB 23471 CTGCGAGTCTTTGTGTGAGGTGTGATCTCTTGGGGGCGGAGGCGTCCCAAC 23412  
QY 570 AlaIleLysGluAspProTrpAsnCysTrpMetCysGlyHisLysGlyThrTyrGlyLe 590  
DB 23411 AGCCATTAAAGAAAGACCTGTGAATGCTGTAATGTGT-----GAGCT 23370  
QY 590 uLeuArgArgArgGluAspTrpProSerArgLeuGlnMetPhePheAlaAsnAsnHisAs 610  
DB 23369 GCTGGCGAAGAGGAGGAGGAGTGGCTTCTGAGCTCCACATGTTCTTTGCCAATACCA 23310  
QY 610 pGlnGluPheAspProProCysValTyrProProValProAlaGluLysArgLysProI 630  
DB 23309 CCAGAAATTTGACCCCCCAAGTTTACCACCTGTGCGAGCTGAGAAAGAGAAACCAT 23250  
QY 630 eArgValLeuSerLeuPheAspGlyIleLeaThrGlyLeuLeuValLeuLysAspLeuI 650  
DB 23249 CACGCGTATCTCTATTGAGGAGGATGCTACAGGCGCTCGGTGCTTAAGAGACTTGG 23190  
QY 650 yIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIleThrValGlyMe 670  
DB 23189 CATCCAAAGTGAACCGTATATGCTTCCGAGGTGTGCGAGCATCATCACCGTGGGCA 23130  
QY 670 tValArgHisGlnGlyLysIleMetTyrValGlyAspValArgSerValThrGlnLysHi 690  
DB 23129 AGTGGCGACCAAGGAAAGATATGTCTCGGGGAGCTCTCAGAGCATACACAGAAACA 23070  
QY 690 sIleGlnGluTrpGlyProPheAspLeuValIleGlyGlySerProCysAsnAspLeuSe 710  
DB 23069 TATCCAGAGAGGAGGCGCATCAATCTGATGTGGGGGCAATCTTGTGAAAGACCTCTC 23010  
QY 710 rIleValAenProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPh 730

```
Db 23009 CATCATCAACCCCTGCCGCAAGGAGCTTTACAAGGCGACATGGCAGCTCTCTTTGAGTT 22950
Qy 730 eTyArgLeuLeuHiaAspAlaArgProLysGluGlyAspArgProPhePheTriPle 750
Db 22949 CTACCACCTCTTCATGATGCGCGGCCCAAGGAGGAGAGGATCGCCCTCTCTTGCTGCT 22890
Qy 750 uPheGluAsnValValAlaMetClyValSerAspLysArgAspLysSerArgPheLeuG1 770
Db 22889 CTTTGAAGAGTGTGGCCATGCGCTTAGTGACACAGAGGACATCTCCGATTTCTTGA 22830
Qy 770 uSerAsnProValMetLysAlaLysGluValSerAlaLysHiaArgAlaArgTyPhe 790
Db 22829 GTCTAAAGCGCTGATGATGATGCTCAAGAAAGTGTCTGCTGCACAGAGGTCCTGTTACTT 22770
Qy 790 eTriPlyAsnLeuProGlyMetAsnArgProLeuAlaSerThrValAsnAspLysLeuG1 810
Db 22769 CTGGGTAAGCTTCTCTGGCATGAACAGGCCATTTGGCATCCACATATGAATGATAAGCTGGA 22710
Qy 810 uLeuGlnGluCysLeuGluHieGlyArgLysAlaLysPheSerLysValArgThrIleTh 830
Db 22709 GCTGCAAGAGTGTCTGCAACATGGCAGAACAGCCAGGTTTCAGCAAAAGTGAGGACCATTTAC 22650
Qy 830 rThrArgSerAsnSerIleLysGlnGlyLysAspGlnHiePheProValPheMetAsnG1 850
Db 22649 CACCAGGTCAACCTCCGTAAGCAGGCGCAAGACACCATTTCCCTGTCTTTCATGAATGA 22590
Qy 850 uLysGluAspLysLeuTriCysThrGluMetGluArgValPheGlyPheProValHieTy 870
Db 22589 GAAGAGACATCTCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 22530
Qy 870 rThrAsp 872
Db 22529 TACAGAA 22523

RESULT 14
AF331857 Homo sapiens DNA cytosine methyltransferase 3 beta (DNMT3B) mRNA PRI 02-JAN-2002
LOCUS complete cds.
DEFINITION AF331857
ACCESSION AF331857
VERSION AF331857.1 GI:18033254
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Ni, J., Pradhan, S. and Roberts, R. J.
Cloning, expression and characterization of human DNMT3 genes
1 (bases 1 to 4335)
2 (bases 1 to 4335)
Ni, J., Pradhan, S. and Roberts, R. J.
Direct Submission
Submitted (22-DEC-2000) New England Biolabs, 32 Tozer Road,
Beverly, MA 01915, USA
Location/Qualifiers
1. .4335
/organism="Homo sapiens"
/db_xref="taxon:9606"
1. .4335
/gene="DNMT3B"
322. .2883
/gene="DNMT3B"
/codon_start=1
/product="DNA cytosine methyltransferase 3 beta"
/db_xref="GI:18033255"
/translation="MKGDTRLNGEEDAGGREDSTLVNAGCSDSSDPPIEAIRTP
EIRGRSSRLSKREVSLLSYTQDLTDGDEGDGSDTTPMKLPRTRTSSESPA
VTRNNNSVSRHRPSPSTGRGRNRHVDSPVEFPATRSRRRATASCTPMP
PSSVLTLDLTDPTDTHGTPOSSSTPYARLAQDSQCGMESPOVEDSGDGSSEYQ
DGKDFGIDLVWGKIKGFSWNPANVVSWKATSKRQMSGMWQWFDGDKFSEVSADK
```

```
LVALGFSQHNLATFNKLVSYRKAMHYHALEKARVRAGKTFPSSPSSDLEDQKPMLE
WAHGGFKPTIGIEGLKNNTQPVNKSQVRAGSRKLESRYENKTRRRRTADSATSQY
CPAPRLKNTCYNNNGKRDGEDOSREOMASDVANNKSLDGLSCGRKNPVSPHPLF
EGGLCOTCDRDLFLFLFYDDQYQSYCTVCCGRELLELLCSNCSRCRCFVCLBLV
GTGTAARLQBPSCMCLPQRCGLVLRKRRKDNVRLQAFFTSDTGLRYEAPKLYPA
IPAARRPRLVLSLFDGLATGVLVLKELGIVKVGKVASVECESTAVGTVKEGNIKY
VNDVNTKNIIEGPPDLVIGGSPCNDLSNVNPAKGLVEGTGLRFFERYHLNYS
RPEGDRDRPFEMFENAVKVGKDRDISPFLNCPNVPMDAIKVSAAHARARVPMGLNP
GMNPVITASKNDKLEODCLEVNRITAKLKVQITTKNSISIKGRKQLFPVYVNSKED
VIMKTELERIFGPPHYHVDVSNMGRAGKQLGSRWSVPVIRHLFAPLKDIFYACE"
BASE COUNT 1045 a 1172 c 1167 g 951 t
ORIGIN
Alignment Scores:
Pred. No.: 5,03e-90 Length: 4335
Score: 2200.00 Matches: 461
Percent Similarity: 60.58% Conservative: 123
Best Local Similarity: 47.82% Mismatches: 265
Query Match: 44.82% Indels: 117
DB: 9 Gaps: 19
US-09-720-086-5 (1-908) x AF331857 (1-4335)
Qy 30 ArgGlyLysGluGlnGluProSerAlaThr---AlaArgLysValGlyArgPro 48
Db 76 CGTGAAGCTCCGAGCGCCCGCCGAGCGGACCGGCTCCCTGGCGTGGGCGAGCG 135
Qy 49 GlyArgLysArgLysHisProValGluSerSerAspThrProLysAspProAlaVal 68
Db 136 GCGGCAACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 195
Qy 69 -----ThrThrLysSerGlnProMetAlaGlnAspSerGlyProSer 82
Db 196 GCATCGCGCGCGGAGATTCGCGAGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCG 255
Qy 83 AspLeuProAsnGlyAspLeuGlyLysArgSerGluProGlnProGlu----- 99
Db 256 CCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 315
Qy 100 -----GluGlySerProAlaAlaGlyGlnLys--- 108
Db 316 GAAAGCATGAAGGAGAGACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAA 375
Qy 109 -----GlyGlyAlaProAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 122
Db 376 GACTCGATCTCGTCAACGCGGCGCTCGACGCGAC---CAGTCTCGACTCGCCCCAATC 432
Qy 123 ---GluAla-----Ser 125
Db 433 CTGAGGCTATCCGACCCCGGAGATCAGAGCGCGAGATCAGCTCGGACTCTCCAAG 492
Qy 126 ArgAlaValGluAsnGlyCysCysValThrLysGluGlyArgGlyAlaSerAlaGlyGlu 145
Db 493 AGGAGGTCTCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 552
Qy 146 GlyLysGluGlnLysGlnThrAsnIle-----GluSerMetLys 158
Db 553 GATGGGATGGCTGTGACACCCCGAGTCATGCCAAAGCTTTCCGGGAAACAGAGCTCGT 612
Qy 159 MetGluGlySerArgGlyArgLeuArgGlyGlyLeuGlyTrpGluSerSerLeuArgGln 178
Db 613 TCAGAAAGCCCGAGCTGTCCGAACCTCGAAATAACACAGTGTCTCCAGCGCGGAGGAC 672
Qy 179 ArgProMetProArgLeuThrPheGlnAlaGlyAspProTyTyIleSerLysArgLys 198
Db 673 AGGCTTCCCGACGTTCCACCC--GAGGCGCGC-----AGGCGC 709
Qy 199 ArgAspGluTrpLeuAla---ArgTTP----- 206
Db 710 GCAACCATGGAGAGTCCCGCTGAGTTCCTCCGCTACCAAGGTCTCTGAGAGCGCGG 769
Qy 207 ---LysArgGluAlaGluLysLysAlaValIleAlaValMetAsnAlaValGluGlu 225
```

Db 770 CAACAGCATCGCAGGAAAGCCATGCGCCCTCCAGCTCTTACCATGAGACC 829  
 QY AsnGlnAlaSerGlyGlnSerGlnValGluGlnAlaSerProProAlaValGlnGln 245  
 Db 830 TCACAGACG-----ACACAGAGACACACATGGGAGCCCCAGAGACAGACATGCC 880  
 QY 246 ProThrAspProAlaSerProThrValAlaThr-----ThrProGluProVal 261  
 Db 881 CTTACG-----CCCGCTAGCCACAGACAGCGAGGGGGGAGCATGAGTCCCGCAG- GTG 936  
 QY 262 GlyGlyAspAlaGlySerPylsAsnAlaThrLysAlaAlaAspAspGluProGlu 281  
 Db 937 GAGCGACAGCTGAGAT-----GAGACAGCTTCAGATGATCAG 975  
 QY 282 AspGlyArgGlyPheGlyLeuGlyValTrpGlyLysLeuArgGlyPheSerTrp 301  
 Db 976 GATGGAGAGAGTTTGGAAATAGGAGACCTGCTGGGGAATCAAGGCTTCTCTGG 1035  
 QY 302 TrpProGlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaGluGlyThr 321  
 Db 1036 TGGCCCGCCATGGTGGTCTTGGAAAGCCACCTCCAAAGCAGAGCTATGCTGGCATG 1095  
 QY 322 ArgTrpValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLysLeuMet 341  
 Db 1096 CGGTGGTCCAGTGGTGGTGGCAGATGGCAGATCTCCAGAGTCTCTGCAACAACTGGTG 1155  
 QY 342 ProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTrpAsnLysGlnProMetTyr 361  
 Db 1156 GCACCTGGGGCTGTTACGCCAGCATTAAATTGGCCACTTCAATAGCTGCTCTAT 1215  
 QY 362 ArgLysAlaIleTrpGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPhePro 381  
 Db 1216 CGAAAGACCATGATACATGCTGCTGGAGAAAGCTAGGGTGAGTGCAAGACCTTCCC 1275  
 QY 382 AlaCysHisAspSerAspGluSerAspSerGlyLysAlaValGluValGlnAsnLysGln 401  
 Db 1276 -----AGCAGCCCTGGAGATCATTTGGAGGACCAAGCTGAAACCC 1314  
 QY 402 MetIleGluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGluProPro 421  
 Db 1315 ATGTGGAGTGGGCCCAAGGGGGCTTCAAGCCCACTGGAGTCAAGGGCTTCAACCCAAC 1374  
 QY 422 GluGlu-----Glu 424  
 Db 1375 AACACGACCAACGCTTAACTAACTCGAAGTGCCTGCTGAGGACGACTAGAAATTTGAA 1434  
 QY 425 LysAsnProTyrLysGluValTyrThrAspMetTrpValGluProGluAlaIle--- 443  
 Db 1435 TCAGAGAAATACGAAACAAAGACTCGAAGACGCAAGCTGACACTGACCCACTTGAC 1494  
 QY 444 TyrAlaProProProAlaLysLysLysProArgLysSerThrThrGluLysProLysVal 463  
 Db 1495 TACGCCCCCGACCCAAAGCGCTTCAAGCAAAATTGCTATTAACAAGCGCAAGACCGAGG 1554  
 QY 464 LysGluIleLeuAspGlyArgTrpArgGluArgLeuValTyrGluValArgGlnLysCys 483  
 Db 1555 GATCAA-----GATCAGAGCCGAGAACAAATGGCTTCAAGTGTGCCAACAACAAG 1605  
 QY 484 ArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThrLeuGlnHisPro 503  
 Db 1606 AGCAGCCTGGAAATGGCTGTTTGTCTTGGCAGAAACCCCGTCTCTTCCACCT 1665  
 QY 504 LeuPheIleGlyLysMetCysGlnAsnLysValAsnLysPheLeuGluCysAlaTyrGln 523  
 Db 1666 CTTCTTGAAGGGGGCTCTGCAACATGCGGGATCGCTTCTTGAAGCTTTTAAATG 1725  
 QY 524 TyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyLysArgGluValLeu 543  
 Db 1726 TATATATACATGCTATCACTTTACTGCACTGTGTCTCTGAGGGCCGAGACTGCTG 1785  
 QY 544 MetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysValAspLeuLeuVal 563  
 Db 1786 CTTTGACAGACACAGAGCTGCTCGGAGTCTTGTGTGAGTGCCTGAGAGTCTGCTG 1845

QY 564 GlyProGlyValAlaGlnAlaIleLysGluAspProTrpAsnLysCysTyrMetCysGly 583  
 Db 1846 GGCACAGCAGCAGCGGCCAGGCCAAAGCTTCAGAGCCCTGAGAGCTTATCATGTGTCTC 1905  
 QY 584 HisLysGlyThrTyrGlyLeuLeuArgArgGluAspTrpProSerTrpGluGlnMet 603  
 Db 1906 CCGCAGCGGTGCATGGCTCTGGCGCCCGGAGAGCATGAGACGTGGCTGCGAGGCC 1965  
 QY 604 PhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyrProProValPro 623  
 Db 1966 TTTCTTACCAAGTACACGCGGCTTGAATATGAAGCCCCCAAGCTGATCCCTGCACTTCCC 2025  
 QY 624 AlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIleAlaThrGlyLeu 643  
 Db 2026 GCAGCCCGAAAGGCGCCCATTTGAGTCCGTGATCTTTGATGGAGTCCGACAGGCTAC 2085  
 QY 644 LeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSerGluValCysGlu 663  
 Db 2086 CTAGTCTTCAAGAGTGGGATGAAGTGAAGAAAGTGGCCCATTTGACTTGGTATGGCGGA 2145  
 QY 664 AspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyrValGlyAspVal 683  
 Db 2146 GAGTCAATGCTGTTGGAACGTGAAGCAGAGGGGAATATCAAAATGCTGAACGACGTG 2205  
 QY 684 ArgSerValThrGlnLysHisIleGlnGluTrpGlyProPheAspLeuValIleGlyGly 703  
 Db 2206 AGGAACATCAAAAGAAATAATTAAGAAATGGGCCCATTTGACTTGGTATGGCGGA 2265  
 QY 704 SerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeuTyrGluGlyThr 723  
 Db 2266 AGCCCATGCAACCATCTTCAATGTGAATCCAGCCAGAGAAAGCCCTGTATGAGGGTACA 2325  
 QY 724 GlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgProLysGluGlyAsp 743  
 Db 2326 GCGCGCTCTCTTCAATTTTACACCTGCTGAATTAAGTCTACAGCCCAAGGGGGTAT 2385  
 QY 744 AspArgProPhePheTrpLeuPheGluAsnValValAlaMetGlyValSerAspLysArg 763  
 Db 2386 GACCGGCGCTTCTTCTGATGTTGAATGTTGTAGCATGAAGTTGGCGCAAGAGG 2445  
 QY 764 AspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLysGluValSerAla 783  
 Db 2446 GACATCTCAGCGTCTCGAGGTATATCCAGTATGATGATGATCCATCAAGTTTCTGCT 2505  
 QY 784 AlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyLysAsnArgProLeuAlaSer 803  
 Db 2506 GCTCAGAGGCGCGATATCTTGGGGCAACTACCGGAGTGAACAGGCCCGATGAC 2565  
 QY 804 ThrValAsnAspLysLeuGluGlnLysCysLeuGluHisGlyArgIleAlaLysPhe 823  
 Db 2566 TCAGAAATATATACTCGAGCTGCAAGACTGCTTGGATACATAGATAGCCAAAGTTA 2625  
 QY 824 SerLysValArgThrIleThrTrpArgSerAsnSerIleLysGlnGlyLysAspGlnHis 843  
 Db 2626 AAGAAATACAGCAATTAACCAACCAAGTGAATCTGATCAACAGGGAAGAAACCACTT 2685  
 QY 844 PheProValPheMetLeuGlnLysGluAspIleLeuTrpCysThrGluMetGluArgVal 863  
 Db 2686 TTCCCTGTTGTCATGATGCAAGAAAGATTTTGTGGCACTGAGCTCGAAAGATC 2745  
 QY 864 PheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeuAlaArgGlnArg 883  
 Db 2746 TTGGCTTCTCTGCACTACAGACGTGTCCAAATGGGCCGTGGTCCCGCCAGAG 2805  
 QY 884 LeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAlaProLeuLysGlu 903  
 Db 2806 CTGTGGGAAGTCCCTGAGACGTGCGCTGATCCGACACTTTCGCGCCCTTGAAGAGAC 2865  
 QY 904 TyrPheAlaCys 907  
 Db 2866 TACTTTCATGT 2877

```

RESULT 15
AF151973      4223 bp      mRNA      linear      ROD 08-JUN-2000
LOCUS        Mus musculus DNA cytosine-specific methyltransferase isoform 5
DEFINITION  AF151973      mRNA, complete cds.
ACCESSION   AF151973
VERSION     AF151973.1  GI:8347127
KEYWORDS
SOURCE      Mus musculus.
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE   1 (bases 1 to 4223)
AUTHORS     Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
            Shen,Y.
TITLE       Cloning of full-length Dnmt3b cDNA and its alternative splicing
            isoforms in mouse embryonic tissue
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 4223)
AUTHORS     Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
            Shen,Y.
TITLE       Direct submission
JOURNAL     Submitted (17-MAY-1999) Department of Biochemistry and Molecular
            Biology, Institution of Basic Medical Sciences, Chinese Academy of
            Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,
            P.R.China
FEATURES
            Location/Qualifiers
            1..4223
             /organism="Mus musculus"
             /strain="KM"
             /db_xref="taxon:10090"
             /dev_stage="8-9 day old embryo"
            1..4223
             /gene="Dnmt3b"
             /gene="Dnmt3b"
             /gene="Dnmt3b"
             /functions="methylates cytosine in DNA"
             /note="alternatively spliced"
             /codon_start=1
             /product="DNA cytosine-specific methyltransferase isoform
             5"
             /protein_id="AAF74519.1"
             /db_xref="GI:8347128"
             /translation="MKGDSRLHNEEGASGYEEIIIVNGNFSQSDTKDAPPPVLE
             AICTPEVCTPETRRSSRLSKREVSLNLYTQDMDGDDRDDEVDGNGSDILMPK
             LTRETKDTRTSPSPASVTRSHNGTSLSLQRQASPRITRQGRHHVQVYVPEFPATR
             SRRRRSSASTPWSPASVDEEVEVTPKSVSTPSVLDSDQDQEGMDTQVDAESIY
             GDSYEQDDKEFGIGDLVWGKIKGFSPWPAVMSWKATSKRQAMPGRWVOWFGDKF
             SEISADKLVALGLFSQHFNLATFNKLVSRKAWYHLEKARVAGTKTPSSPESGLED
             QLKPMLEWAGHGFKPTGIEGLKPNKKQPVNKSXVRESDSRNLEPRRENKSRRTTN
             DSAASEPPPKRLKNTSYGCKDRGEDESRERMASEVTNNKNLEDRCLSCCKNPVS
             FPLVFGGLCQSCDRDELFLFYMYDEDCYQSYCTVCCGREGILLCSCTSCFCVEG
             LEVLVGAETADAKLOEPWSCVYKLPORCHGVLRKDKNMELQDFPTDPLSEFER
             PKLYPAIAKRPRIPLSLPDIATGLVILKELGKVEKYIASEVCAESIATVTKH
             EGQIKYVNDVRKITKENIEEMQPDVLVIGGSPCNILSNVNPARKGLYEGTGRLEFEFY
             HLLAYTEPKEDNRPFPMFNVAMKNDKOISRLFACNPFVMDIAIKVSAHKARY
             FWNGLPMPNRPWASNDKLELODCLERSRTAKLKKRVOTITKASINSIRQGNQLFPVY
             MNGKDDVLWCTELERIFGPFPHYTDVSNMGRGKQLKLGWSVVPVIRHLFAPLKDYP
             ACE"
BASE COUNT  1051 a 1084 c 1147 g 941 t
ORIGIN
Alignment Scores:
Pred. No.:      5,72e-90      Length:      4223
Score:          2198.50      Matches:      449
Percent Similarity: 60.42%      Mismatches: 131
Best Local Similarity: 46.77%      Indels:      143
Query Match:    44.79%      Gaps:        15
DB:
US-09-720-086-5 (1-908) x AF151973 (1-4223)
QY 44 LysValGlyArgProGlyArgLysArgLys-----53

```

```

Db 130 CGCAGCGGCGCTCCAGTCCGCGCAGCGCGGGTTAAGCGGCCCAAGTAAACGTAGCCGAC 189
QY 54 -----HisProProValGluSerSerArgThrProLysAspProAlaValThrThrLys 71
Db 190 CGATCGGCGCGGAGATTTCGGAACCCGACACTCCGCGCGCGCGCGCGGCGGACCCGC 249
QY 72 -----SerGlnProMetAlaGlnAspSerGlyProSerAsp 83
Db 250 GGGCGCATCGGCGCGCGCGCTACGCCACCTCA-CGACAGGAACAACATAAGGGAGAC 308
QY 84 -----LeuLeuProAsn 87
Db 309 AGCAGACATCTGAATGAAGAAGAGGGTGCACGGGTATGAGGAGTGCAATTCGTTAAT 368
QY 88 GlyAspLeuGluLysArgSerGluProGlnProGluGluGlySerProAlaAlaGlyGln 107
Db 369 GGGAACTTTCAGTGACCGAGTCTCTCAGACACCAAGAGTGTCTCCCTCACC- 416
QY 108 LysGlyGlyAlaProAlaGluGlyGluGlyThrGluThrProProGluAlaSerArgAla 127
Db 417 -----CCAGTCTTGGAGGCAATCTGC 437
QY 128 ValGluAsnGlyCysCysValThrLysGluGlyArgGlyAlaSerAla----- 143
Db 438 ACAGAGCCAGTCTGCACACCCAGACAGCCAGGCGCGAGTCAAGTCCCGGCTGTCTAAG 497
QY 144 -----GlyGluGly----- 146
Db 498 AGGGAGTCTCCAGCTTCTGAATTACACGAGGACATGACAGGAGATGGAGACAGAT 557
QY 147 -----LysGluGln 149
Db 558 GATGAAGTAGATGATGGAAATGGCTCTGATATCTTAATGCCAAAGCTCACCGTGAGACC 617
QY 150 LysGlnThrAsnIleGluSerMetLysMetGluGlySerArgGlyArgLysArgGlyGly 169
Db 618 AAGGACACCGAGCGCGCTCT-----GAAAGCCGCGCTGTCCGAACCCGACATAGC 668
QY 170 LeuGlyTrpGluSerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGly 189
Db 669 AATGGGACCTCCAGCTTGGAGAGGCAAGAGGCTCCGCCAGATCACCAGGTCGGCAG 728
QY 190 AspProTyrTyrIleSerLysArgLysArgGluTrpLeuAlaArgTrpLysArgGlu 209
Db 729 GGGCCGCCACCATGTG-----CAGAGTACCTCTGGAGTTTCCGGTACC 773
QY 210 AlaGluLysLysAlaLysValIleAlaValMetAsnAlaValGluGluAsnGlnAlaSer 229
Db 774 AGGTCTCGAGACCTCGCATCGCTTCAGCAGCAGCCCATGGTCATCCCCCTGCCAGC 833
QY 230 GlyGluSerGlnLysValGluGluAlaSerProProAlaValGlnGlnProThrAspPro 249
Db 834 GTCGACTTC-----ATGGAAGAAGTGCACACTAAGAGCGGTAGTACCCCA----- 878
QY 250 AlaSerProThrValAlaThrThrProGluProValGlyGlyAspAlaGlyAspLysAsn 269
Db 879 --TCAGTTGACTGTGAGCCAGGATGGAGATCAGGAGGTATGGATACCACAGAGTGGAT 935
QY 270 AlaThrLysAlaAlaAspAspGluProGluTrpGluAspGlyArgGlyPheGlyIleGly 289
Db 936 GCAGAGACATATATATGGACACAGCACAGATATCAGGATGATAAAGAGTTTGAATAGT 995
QY 290 GluLeuValTrpGlyLysLeuArgGlyPheSerTrpTrpProGlyArgIleValSerTrp 309
Db 996 GACCTCGTGTGGGGAAGATCAAGGGCTTCTCTGTTGGCTGCCATGGTGGTGTCTTGG 1055
QY 310 TrpMetThrGlyArgSerArgAlaAlaGluGlyThrArgTrpValMetTrpPheGlyAsp 329
Db 1056 AAAGCCACCTCCAAAGCGACAGCCCATGCCCGAATGCCGTGGGTGAGTGGTTTGTGAT 1115
QY 330 GlyLysPheSerValValCysValGluLysLeuMetProLeuSerSerPheCysSerAla 349

```

Db 1116 GCGAAGTTTCTGAGATCTGCTGACAACTGCGCTGGGGCTTCCAGCCAGCAC 1175  
Qy 350 PheHisGlnAlaThrTyrAsnLysGlnProMetTyrArgLysAlaLeuTyrGluValLeu 359  
Db 1176 TTAAATCTGCTACCTTCATTAAGCTGGTTTCTTATAGAAAGGCATGTACCACTCTG 1235  
Qy 370 GlnValAlaSerSerArgAlaGlyLysLeuPheProAlaCysHisAspSerArgLysSer 389  
Db 1236 GAAAGAGCCGAGGTTCCAGCTGGCCAGACCTTC-----TTCAGC 1274  
Qy 390 AspSerGlyLysAlaValGluValGlnAsnLysGlnMetIleGluTyrAlaLeuGlyLys 409  
Db 1275 AGTCCTGGAGAGTCACCTGGAGACCACTGAGCCATGCTGGAGGCGCCAGGTGGC 1334  
Qy 410 PheGlnProSerGlyProLysGlyLeuGluPro----- 420  
Db 1335 TTCAGGCTTCTGGATTCGAGGGGCTCAAAACCAAGAGACCAAGCGTTAATAAG 1394  
Qy 421 -----ProGluGluLysAsnProTyr 428  
Db 1395 TCGAAGGTGCGTCTTCAGACAGTAGAGAACTTAGAACCCGAGACCGAGAACAAAGT 1454  
Qy 429 LysGluValTyrThrAspMetTyrPvalGluProGlnAlaAlaTyrAlaProPro 448  
Db 1455 CGAAGACGCAACCAAT-----GACTCTGCTGCTGTAGTCCCCCA 1499  
Qy 449 ProAlaLysProArgLysSerThrThrGluLysProLysValLysGluIleLeuAsp 468  
Db 1500 CCCAAGGCTCTCAAGACAAATAGCTATGGCGGAGAGACCGAGGGAG-----GAT 1550  
Qy 469 GluArgThrArgGluArgValTyrGluValArgLysLysCysArgAsnIleGluAsp 488  
Db 1551 GAGGAGAGCCGAGAACGAGTGGCTTCTGAAGTCCACAAACAAAGGCAATCTGGAAGAC 1610  
Qy 489 IleCysIleSerCysGlySerLeuAsnValThrLeuGluHisProLeuPheIleGly 508  
Db 1611 CGCTGTTGCTGTGGAAAGAACCTGTGTCTTCCACCCCTTTGAGGGTGG 1670  
Qy 509 MetCysGlnAsnCysLysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAspArgLys 528  
Db 1671 CTCTGTAGAGTTCGCGGATCGCTTCTTAGAGCTCTTTCATGTATGATGAGAGCGC 1730  
Qy 529 TyrGlnSerTyrCysThrIleCysCysGlyValArgGluValLeuMetCysGlyAsnAsn 548  
Db 1731 TATCAGTCTACTGACACCGTGTGCTGTGAGGGCGCTGAACGTGCTGTGAGTAACACA 1790  
Qy 549 AsnCysCysArgCysPheCysValGluCysValAspLeuLeuValGlyProGlyAlaAla 568  
Db 1791 AGCTGCTGCAGATGCTTCTGTGTGAGTGTCTGAGAGTGTGTGGGCGCAGGCACAGCT 1850  
Qy 569 GlnAlaAlaIleLysGluAspProTyrAsnProTyrMetCysGlyHisLysGlyThrTyr 588  
Db 1851 GAGGATGCCAAGCTGCAGAGACCTTGAAGCTATATGTGCTTCCCTCAGCGCTGCCAT 1910  
Qy 589 GlyLeuLeuArgArgLysAspTyrProSerArgLeuGlnMetPhePheAlaAsnAsn 608  
Db 1911 GGGGTCTCCGAGCAGGAAAGATGGAACTGCGCTGCAGAACTTCTTCACTACATGAT 1970  
Qy 609 HisAsp---GlnGluPheAspProProLysValTyrProProValProAlaGluLysArg 627  
Db 1971 CTTGACCTGGAAAGATTGAGCCCAAGTTGTACCAGCAATTCCTGCAGCCAAAAGG 2030  
Qy 628 LysProIleArgValLeuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeuLys 647  
Db 2031 AGGCCCATTTAGTCTGCTCTGTGTGTGATGAAATGGCAACGGGGTACTGTGCTCAAG 2090  
Qy 648 AspLeuGlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIleThr 667  
Db 2091 GAGTTGGGTATTAAAGTGAAGAACTGCTCCGAGAGCTGTGCAAGATCCATGCT 2150  
Qy 668 ValGlyMetValArgHisGlnGlyLysIleMetCysValGlyAspValArgSerValThr 687  
Db 2151 GTGGGAAGCTTTAAGCATGAAGGCCAGATCAAAATATGTCAATGAGTCCGGAAATCACC 2210

Qy 688 GlnLysHisIleGlnGluTyrGlyProPheAspLeuValIleGlyCysSerProCysAsn 707  
Db 2211 AAGAAAAATATTGAAGAGTGGGGCCCGTTTCGATCTGTGATGTGTGGAGGCCCATCAAT 2270  
Qy 708 AspLeuSerIleValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhe 727  
Db 2271 GATCTCTTAACGTCAATCTTCCGCAAGAGTTTATATGAGGGGCACAGAAAGGCTTTC 2330  
Qy 728 PheGluPheTyrArgLeuLeuHisAspAlaArgProLysGluGlyAspAspArgProPhe 747  
Db 2331 TTGAGATTTTACACTGCTGTAATTATACCCGCCCAAGAGGGCGACACAGCTTCATTC 2390  
Qy 748 PheTyrLeuPheGlnAsnValAlaAlaMetGlyValSerAspIleArgAspIleSerArg 767  
Db 2391 TTCTGATGTTCCAGAAATGTTGGCCATGAAGAAAGTGAATGAAAGCAAGAAACATCTCAAG 2450  
Qy 768 PheLeuGluSerAsnProValMetIleAspAlaLysGluValSerIleAlaHisArgAla 787  
Db 2451 TTCTTGACATGTAAACCACTGATGATGATGCCATCCATCAAGGTGTCTCTCACAGGGCC 2510  
Qy 788 ArgTyrPheTyrGlyAsnLeuProGlyMetAsnArgProLeuAlaSerThrValAsnAsp 807  
Db 2511 CGGTACTTCTGGGGTAACTTACCCGGAATGAACAGGCCCTGTATGCTTCAAGAAATGAT 2570  
Qy 808 LysLeuGluLeuGlnLysLeuGluCysLeuGluHisGlyArgIleAlaLysPheSerLysValArg 827  
Db 2571 AAGCTGAGCTGACAGAGCTGCTGAGATTCAATGAGACAGCAAGAAAGTTAAAGAAAGTCAG 2630  
Qy 828 ThrIleThrThrArgSerAsnSerIleLysGlnGlyLysAspGlnHisPheProValPhe 847  
Db 2631 ACAATTAACCAACCAAGTCGAATCTTCATCAGACAGGGCAAAACCAAGCTTTCCTGTATGC 2690  
Qy 848 MetAsnGluLysGluAspIleLeuTyrCysThrGluMetGluArgValPheGlyPhePro 867  
Db 2691 ATGAATGGCAGAGACAGATGTTGTGTGCTGACTGAAAGATCTTCGGCTTCCCT 2750  
Qy 868 ValHisTyrThrAspValSerAsnMetSerArgLeuAlaArgGlnArgLeuLeuGlyArg 887  
Db 2751 GTCACATACAGGACGAGTTCATCAACATGGGCGCGGCGCGCTCAGAAAGCTGCTGGCAGG 2810  
Qy 888 SerTyrSerValProValIleArgHisLeuPheAlaProLeuLysGluTyrPheAlaCys 907  
Db 2811 TCCTGAGGTGACCGGTATCAGACACCTGTTGCCCCCTTGAAGAGACTACTTGTGCTGT 2870

Search completed: November 22, 2002, 10:51:23  
Job time : 5545.7 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2002, 08:17:42 ; Search time 530.945 Seconds

(without alignments)  
1780.325 Million cell updates/sec

Title: US-09-720-086-1

Perfect score: 4192

Sequence: 1 gaattccggcctcgtcgccg.....taaaaaacatacaaatgct 4192

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0  
2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: N\_Geneseq\_101002.\*  
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4181.4	99.7	4192	21	AA237095
2	2871.8	68.5	4416	21	AA237097
3	1814	2938	24	AB190391	Human polynucleoti
4	757.8	18.1	4195	21	AA237096
5	749.8	17.9	4145	21	AA237098
6	559	13.3	709	24	AB057540
7	547.6	13.1	1060	22	AAH33860
8	482.8	11.5	3017	22	AAH14009
9	413	9.9	622	22	AA100271

10	341	8.1	711	23	AA586165	DNA encoding novel
11	324.6	7.7	2077	22	AA580537	Receptor #25 parti
12	177.4	4.2	301	16	AA721884	Human gene signatu
13	171.4	4.1	546	22	ABA20255	Human nervous syst
14	169.8	4.1	2951	23	AA586166	DNA encoding novel
15	167	4.0	283	22	ABA13731	Human nervous syst
16	165.8	4.0	411	22	AA185785	Human polynucleoti
17	141	3.4	1705	22	AA582964	Human DNA methyltr
18	141	3.4	3002	21	AA582964	Human ORFX ORF3010
19	102.6	2.4	762	22	AAH05520	Human cDNA clone (
20	97.8	2.3	273	24	AAH97389	Gene #3887 used to
21	97.8	2.3	273	24	AB164228	Stomach cancer rel
22	90.2	2.2	573	24	AB182861	Human ovarian canc
23	78	1.9	237	24	AB177340	Human breast cell
24	77.6	1.9	448	22	ABA43755	Human breast cell
25	77.6	1.9	448	22	ABA54712	Human breast cell
26	77.6	1.9	448	22	ABA23961	Human foetal liver
27	77.6	1.9	448	22	AAK02890	Probe #2427 for ge
28	77.6	1.9	448	22	AAK27928	Human brain expres
29	77.6	1.9	448	22	AAK112506	Human bone marrow
30	77.6	1.9	448	22	AA133860	Probe #2546 used t
31	77.6	1.9	448	22	AA102414	Probe #2405 used t
32	77.6	1.9	448	24	AB502382	Human genome-deriv
33	71.2	1.7	430	22	AA185778	Human polynucleoti
34	65	1.6	65	24	ABN51598	Mouse spliced tran
35	62	1.5	384	22	ABA49011	Human foetal liver
36	62	1.5	384	22	ABA66923	Probe #12475 for g
37	62	1.5	384	22	AAK15369	Human brain expres
38	62	1.5	384	22	AAK41092	Human bone marrow
39	62	1.5	384	22	AA121858	Probe #11791 for g
40	62	1.5	384	22	AA147141	Probe #15827 used
41	62	1.5	384	22	AA107545	Probe #7536 used t
42	62	1.5	384	22	AB515085	Human genome-deriv
43	62	1.4	3973	24	AB132171	Human immune syste
44	58.4	1.4	3973	24	AB132171	Human polynucleoti
45	58	1.4	398	22	AA187804	Human polynucleoti

## ALIGNMENTS

RESULT 1	AA237095	DNA; 4192 BP.
ID	AA237095	standard; DNA; 4192 BP.
AC	AA237095;	
XX		
DT	27-MAR-2000	(first entry)
XX		
DE	DNA encoding de novo DNA cytosine methyltransferase Dnmt3a.	
XX		
KW	De novo DNA cytosine methyltransferase; Dnmt3a; neoplastic disorder;	
KW	carcinoma; sarcoma; leukaemia; DNA methylation; ss.	
XX		
OS	Mus sp.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	217..2943
FT	/*tag= a	
FT	/product= "de novo DNA cytosine methyltransferase	
XX		Dnmt3a"
XX		
XX	W09967397-A1.	
XX		
PD	29-DEC-1999.	
XX		
PF	25-JUN-1999;	99WO-US14373.
XX		
PR	25-JUN-1998;	98US-0090906.
PR	24-JUL-1998;	98US-0093993.
XX		
XX		
PA	(GEHO ) GEN HOSPITAL CORP.	

PI Li E, Okano M, Xie S;  
XX WPI: 2000-106298/09.  
DR P-FSD8; AAY54055.  
XX  
PT New mouse and human polypeptides, useful to treat and diagnose  
PT neoplastic disorders e.g. carcinomas, sarcomas and leukemias -  
XX  
XX Claim 8; Fig 1A; 114pp; English.  
XX  
CC The present sequence encodes a murine de novo DNA cytosine  
CC methyltransferase designated Dnm3a. The polypeptides can be administered  
CC therapeutically, especially by expressing encoding polynucleotides, to  
CC treat diseases associated with DNA cytosine methyltransferase, such as  
CC neoplastic disorders e.g. carcinomas, sarcomas and leukemias. They can  
CC be used to diagnose, or determine susceptibility to neoplastic disorders,  
CC by assaying for polypeptide expression levels in mammalian cells/body  
CC fluids. They are useful to screen for compounds inhibiting/activating the  
CC polypeptide. The polypeptides can also be used for in vitro de novo  
CC methylation of DNA. Such in vitro methylation may be used to direct or  
CC regulate DNA expression in biological systems, e.g. recombinant DNA  
CC methylated in vitro may be introduced into a cell/organism to increase  
CC or decrease expression of a desired polypeptide for which the native DNA  
CC is under-methylated or not methylated. The polypeptides can also be  
CC used to produce antibodies which are useful to detect and purify the  
CC polypeptide or therapeutically e.g. to treat neoplastic disorders. The  
CC polynucleotides are useful to produce probes and primers which are  
CC useful diagnostically.  
XX  
SQ Sequence 4192 BP; 1095 A; 1070 C; 1182 G; 844 T; 1 other;  
Query Match 99.7%; Score 4181.4; DB 21; Length 4192;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 4186; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Qy 1 GAATTCGGGCTCTCTCCGGGCGCCCGACCCCGGGCCACACGCGAGCGCGCTGAA 60  
Db 1 GAATTCGGGCTCTCTCCGGGCGCCCGACCCCGGGCCACACGCGAGCGCGCTGAA 60  
Qy 61 GCCCAGCGCTGAGGCTGCACTTTTCCGAGGGCTTGACATCAGGGTCTATGTTAAAGTCTT 120  
Db 61 GCCCAGCGCTGAGGCTGCACTTTTCCGAGGGCTTGACATCAGGGTCTATGTTAAAGTCTT 120  
Qy 121 AGCTCTTGTCTTACAAAGACCACGCAATTCCTCTCTGAAGCCCTCGCAGCCCCACAGCG 180  
Db 121 AGCTCTTGTCTTACAAAGACCACGCAATTCCTCTCTGAAGCCCTCGCAGCCCCACAGCG 180  
Qy 181 CCCTCGAGCCCGCAGCTCGCCTACTGCCCAGCATGCTCCAGCGGCCCGGGGAC 240  
Db 181 CCCTCGAGCCCGCAGCTCGCCTACTGCCCAGCATGCTCCAGCGGCCCGGGGAC 240  
Qy 241 ACCAGAGCTCTCTCTGAGCGGAGGATGATCGAAGGAAGGAGGAGGAGGAG 300  
Db 241 ACCAGAGCTCTCTCTGAGCGGAGGATGATCGAAGGAAGGAGGAGGAGGAGGAG 300  
Qy 301 AACCGTGGCAAGGAGCGCCAGAGCCCGAGCCCGCCAGCGCCCGGAGGAGGCGCT 360  
Db 301 AACCGTGGCAAGGAGCGCCAGAGCCCGAGCCCGCCAGCGCCCGGAGGAGGCGCT 360  
Qy 361 GGCCGGAAGCGCAAGCACCCACCGGTGGAAAGCAGTGACACCCCGGAGGAGGAGGAGG 420  
Db 361 GGCCGGAAGCGCAAGCACCCACCGGTGGAAAGCAGTGACACCCCGGAGGAGGAGGAGG 420  
Qy 421 ACCACCAAGTCTCAGCCCATGGCCAGGACTCTGGCCCTCAGATCTGCTACCCCAATGGA 480  
Db 421 ACCACCAAGTCTCAGCCCATGGCCAGGACTCTGGCCCTCAGATCTGCTACCCCAATGGA 480  
Qy 481 GACTTGGAGAGCGGAGTGAACCCCAACTCAGGAGGGAGCCAGCTCGAGGGCAGAGG 540  
Db 481 GACTTGGAGAGCGGAGTGAACCCCAACTCAGGAGGGAGCCAGCTCGAGGGCAGAGG 540  
Qy 541 GGTGGGCGCCAGCTCAAGGAGGAGGAACTGAGACCCACAGAGCCCTCCAGAGCTGTG 600  
Db 541 GGTGGGCGCCAGCTCAAGGAGGAGGAACTGAGAGTGGCCAGAGAGTGCAGNAACATCGAGGC 1680

Db 541 GGTGGGCGCCAGCTGAAGGAGAGGGAACGTGAGACCCACAGAAAGCTCCAGAGCTGTG 600  
Qy 601 GAGAAATGCTGCTGTGTGACCAAGGAAGCCCTGTGAGAGCTCTGTGAGAGAGGCAAGAA 660  
Db 601 GAGAAATGCTGCTGTGTGACCAAGGAAGCCCTGTGAGAGCTCTGTGAGAGAGGCAAGAA 660  
Qy 661 CAGAGAGCAGCAACATCGAATCGAATCGAATCGAATCGAATCGAATCGAATCGAATCGAAT 720  
Db 661 CAGAGAGCAGCAACATCGAATCGAATCGAATCGAATCGAATCGAATCGAATCGAATCGAAT 720  
Qy 721 GGTGGGCTGGGAGTGCAGCTCCGTGAGCGAACCCATGCCAAGACTCACCTTCCAGGCA 780  
Db 721 GGTGGGCTGGGAGTGCAGCTCCGTGAGCGAACCCATGCCAAGACTCACCTTCCAGGCA 780  
Qy 781 GGGGACCCCTACTACATCAGCAAAACGGAAACGGGATGAGTGGCTGGCACCTTGGAAAGG 840  
Db 781 GGGGACCCCTACTACATCAGCAAAACGGAAACGGGATGAGTGGCTGGCACCTTGGAAAGG 840  
Qy 841 GAGGCTGAGAAAGAAAGCAAGGTAATTCAGTAATGAATGCTGTGGAAAGAAACCAAGGCC 900  
Db 841 GATGCTGAGAAAGAAAGCAAGGTAATTCAGTAATGAATGCTGTGGAAAGAAACCAAGGCC 900  
Qy 901 TCTGGAGAGTCTCAGAAAGGTGGAGAGGCGACCCCTCTGCTGTGCGAGAGCCACGGAC 960  
Db 901 TCTGGAGAGTCTCAGAAAGGTGGAGAGGCGACCCCTCTGCTGTGCGAGAGCCACGGAC 960  
Qy 961 CTTGCTTCTCCGACTGTGGCCACACCCCTGAGCCAGTGAAGGGGATGCTGGGAGCAAG 1020  
Db 961 CTTGCTTCTCCGACTGTGGCCACACCCCTGAGCCAGTGAAGGGGATGCTGGGAGCAAG 1020  
Qy 1021 AATGCTACAAAGCAGCGAGATGAGCTGATGATGAGGATGCGCCGGGCTTTGGCAT 1080  
Db 1021 AATGCTACAAAGCAGCGAGATGAGCTGATGATGAGGATGCGCCGGGCTTTGGCAT 1080  
Qy 1081 GGAGAGCTGGTGTGGGGAACCTTCCGGGCTTCTCTGCTGGCCAGGCCGAAATGTGTCT 1140  
Db 1081 GGAGAGCTGGTGTGGGGAACCTTCCGGGCTTCTCTGCTGGCCAGGCCGAAATGTGTCT 1140  
Qy 1141 TGTGATGACAGCCCGGAGCCGAGCAGCTGGAAGGCACTCGCTGGGCTCATGTGTTCCGA 1200  
Db 1141 TGTGATGACAGCCCGGAGCCGAGCAGCTGGAAGGCACTCGCTGGGCTCATGTGTTCCGA 1200  
Qy 1201 GATGGCAAGTTCCTCAGTGGTGTGTGGGAAGCTCATGTGCGGCTGAGCTCTCTGCGAGT 1260  
Db 1201 GATGGCAAGTTCCTCAGTGGTGTGTGGGAAGCTCATGTGCGGCTGAGCTCTCTGCGAGT 1260  
Qy 1261 GCATTCACAGGCGCACTACAAAGCAGCCCATGTATCCGAAAGCCATCTACGAAGTC 1320  
Db 1261 GCATTCACAGGCGCACTACAAAGCAGCCCATGTATCCGAAAGCCATCTACGAAGTC 1320  
Qy 1321 CTCAGGTTGGCCAGCGCCGCGGGAGCTGTTTCCAGCTTGCATGACAGTGTGA 1380  
Db 1321 CTCAGGTTGGCCAGCGCCGCGGGAGCTGTTTCCAGCTTGCATGACAGTGTGA 1380  
Qy 1381 AGTGAAGTGGCAAGGCTGTGGAAGTGCAGAACAGAGAGATGATGATGGGCTTCGGT 1440  
Db 1381 AGTGAAGTGGCAAGGCTGTGGAAGTGCAGAACAGAGAGATGATGATGGGCTTCGGT 1440  
Qy 1441 GGTTCACAGCCCTCGGGTCTTAAAGGCTTGGAGCCACAGAGAGGAGAAATCTCTTAC 1500  
Db 1441 GGTTCACAGCCCTCGGGTCTTAAAGGCTTGGAGCCACAGAGAGGAGAAATCTCTTAC 1500  
Qy 1501 AAGGAAGTTTACACCGACATGTGGGTGGAGCCCTGAAGCAGCTGTACGCCCCACCCCA 1560  
Db 1501 AAGGAAGTTTACACCGACATGTGGGTGGAGCCCTGAAGCAGCTGTACGCCCCACCCCA 1560  
Qy 1561 CCAGCAAGAAACCCAGAAAGAGCACACAGAGAAACCTTAAGGTCAAGGAGATCATTTGAT 1620  
Db 1561 CCAGCAAGAAACCCAGAAAGAGCACACAGAGAAACCTTAAGGTCAAGGAGATCATTTGAT 1620  
Qy 1621 GAGCGCACAGGGAGCGGCTGTGTATGAGGTGCGCCAGAGAGTGCAGAAACATCGAGGAC 1680  
Db 1621 GAGCGCACAGGGAGCGGCTGTGTATGAGGTGCGCCAGAGAGTGCAGAAACATCGAGGAC 1680



QY 1681 ATTTGTATCTCATGTGGAGCCTCATGTCACTCTGAGACAACCACTTTCATTTGAGGC 1740  
DB 1681 ATTGTATCTCATGTGGAGCCTCATGTCACTCTGAGACAACCACTTTCATTTGAGGC 1740  
QY 1741 ATGGCCAGAACTGTAGAACTGCTTCTTGGAGTGTCTTAACAGTATGACAGATGGG 1800  
DB 1741 ATGGCCAGAACTGTAGAACTGCTTCTTGGAGTGTCTTAACAGTATGACAGATGGG 1800  
QY 1801 TACCAGTCTATTGCAACATCTGCTGTGGGGGCGCTGAAGTCTCATGTGTGGGAACAC 1860  
DB 1801 TACCAGTCTATTGCAACATCTGCTGTGGGGGCGCTGAAGTCTCATGTGTGGGAACAC 1860  
QY 1861 AACTGTGACAGTGTCTTTTGTGTGAGTGTGTGATCTCTTGTGGGGCAGAGCTGCT 1920  
DB 1861 AACTGTGACAGTGTCTTTTGTGTGAGTGTGTGATCTCTTGTGGGGCAGAGCTGCT 1920  
QY 1921 CAGGAGCCATTAAAGAAAGACCCCTGGAACCTGCTACATGTGCGGGCATTAAGGCACTTAT 1980  
DB 1921 CAGGAGCCATTAAAGAAAGACCCCTGGAACCTGCTACATGTGCGGGCATTAAGGCACTTAT 1980  
QY 1981 GGGCTGTGCGAAGACCGGAAGACTGCGCTTCTGCACTCCAGATGTTCTTTGCGAATTAAC 2040  
DB 1981 GGGCTGTGCGAAGACCGGAAGACTGCGCTTCTGCACTCCAGATGTTCTTTGCGAATTAAC 2040  
QY 2041 CATGACCAAGAAATTTGACCCGCCAAGGTTTACCACCTGTGCGAGCTGAGAAAGAGAG 2100  
DB 2041 CATGACCAAGAAATTTGACCCGCCAAGGTTTACCACCTGTGCGAGCTGAGAAAGAGAG 2100  
QY 2101 CCCATCCGCTGTCTGTCTCTTCTTGTGATGGATGTCTACAGGCTCTGTGTGTGAGAGAC 2160  
DB 2101 CCCATCCGCTGTCTGTCTCTTCTTGTGATGGATGTCTACAGGCTCTGTGTGTGAGAGAC 2160  
QY 2161 CTGGGACCTCAAGTGAACCGCTACATTTGCTCCGAGGTGTGTGAGACTCCATCAACGGTG 2220  
DB 2161 CTGGGACCTCAAGTGAACCGCTACATTTGCTCCGAGGTGTGTGAGACTCCATCAACGGTG 2220  
QY 2221 GGCATGTGTGGGACCAAGGAAAGATCATGTACGTCCGGGACGTCCGCAAGCTCACACAG 2280  
DB 2221 GGCATGTGTGGGACCAAGGAAAGATCATGTACGTCCGGGACGTCCGCAAGCTCACACAG 2280  
QY 2281 AAGCATATCCAGAGTGGGGCCCATTCGACTGTGTGATTTGAGGCACTCCCTGCAATGAC 2340  
DB 2281 AAGCATATCCAGAGTGGGGCCCATTCGACTGTGTGATTTGAGGCACTCCCTGCAATGAC 2340  
QY 2341 CTCTCAATTTGTCAACCTGCGCGAAGGGACTTATATGAGGTACTGCGCCCTCTTTCTT 2400  
DB 2341 CTCTCAATTTGTCAACCTGCGCGAAGGGACTTATATGAGGTACTGCGCCCTCTTTCTT 2400  
QY 2401 GAGTTCTACCGCTCTCTGATGTGCGGGCCCAAGAGGGAATGATGCGCCCTTCTTC 2460  
DB 2401 GAGTTCTACCGCTCTCTGATGTGCGGGCCCAAGAGGGAATGATGCGCCCTTCTTC 2460  
QY 2461 TGGCTCTTTGAGATGTGTGGCCATGGGCTTATGTAACAAGAGGACATCTGCGCATTT 2520  
DB 2461 TGGCTCTTTGAGATGTGTGGCCATGGGCTTATGTAACAAGAGGACATCTGCGCATTT 2520  
QY 2521 CTGGAGCTTAACCCCGTGAATTTGAGCCAAAGAAATGTCTGTGTGACACAGGGCCCGT 2580  
DB 2521 CTGGAGCTTAACCCCGTGAATTTGAGCCAAAGAAATGTCTGTGTGACACAGGGCCCGT 2580  
QY 2581 TACTTGTGGGTAACTTCTCTGAGATGAACAGGCTTTGGCATCTGATGATGATTAAG 2640  
DB 2581 TACTTGTGGGTAACTTCTCTGAGATGAACAGGCTTTGGCATCTGATGATGATTAAG 2640  
QY 2641 CTGAGCTGCAAGAGTGTCTGAGCAACGCAAGTAATAGCAAGTTCAAGCAAAAGTGAAGACC 2700  
DB 2641 CTGAGCTGCAAGAGTGTCTGAGCAACGCAAGTAATAGCAAGTTCAAGCAAAAGTGAAGACC 2700  
QY 2701 AATAACCAAGGTCAAACTCTATTAAGACAGGGCAAAACACAGCAATTTCCCGCTTTATG 2760  
DB 2701 AATAACCAAGGTCAAACTCTATTAAGACAGGGCAAAACACAGCAATTTCCCGCTTTATG 2760

QY 2761 AACGAGAGGAGACATCTCTGTGTGCACTGAAATGGAAGGCTGTTTGGCTTCCCGTCC 2820  
DB 2761 AACGAGAGGAGACATCTCTGTGTGCACTGAAATGGAAGGCTGTTTGGCTTCCCGTCC 2820  
QY 2821 CACTACACAGACGTCTCCAAATGAGCCGCTTGCGAGGAGACCTGTGGGCCGATTCG 2880  
DB 2821 CACTACACAGACGTCTCCAAATGAGCCGCTTGCGAGGAGACCTGTGGGCCGATTCG 2880  
QY 2881 TGGAGCGTCCGGTCAATCCGCCACTCTTGTGCTCCGCTGGAAGAAATTTTGTGTGTG 2940  
DB 2881 TGGAGCGTCCGGTCAATCCGCCACTCTTGTGCTCCGCTGGAAGAAATTTTGTGTGTG 2940  
QY 2941 TAAAGGACATGSGGGCAAACTGAAGTATGATATATAAAAAAGTTAAACAAACAAAC 3000  
DB 2941 TAAAGGACATGSGGGCAAACTGAAGTATGATATATAAAAAAGTTAAACAAACAAAC 3000  
QY 3001 AAAAAACAAAAACAAATATAACACAAAGAACAGAGACGAGAAAAAGTTACAGCAC 3060  
DB 3001 AAAAAACAAAAACAAATATAACACAAAGAACAGAGACGAGAAAAAGTTACAGCAC 3060  
QY 3061 CAGAGAGAAAAAGAAATTTAAAGCAAAACAGAGAGGAAAAAGCGCGGAGGCTTGGC 3120  
DB 3061 CAGAGAGAAAAAGAAATTTAAAGCAAAACAGAGAGGAAAAAGCGCGGAGGCTTGGC 3120  
QY 3121 CTTCGAAAAAGGTTGACATCATCTCTGATTTTCAATGTATTAACCTTCACTATCTA 3180  
DB 3121 CTTCGAAAAAGGTTGACATCATCTCTGATTTTCAATGTATTAACCTTCACTATCTA 3180  
QY 3181 AAAAAACAAATAGCCCTCTCCCTTCTTCCCTCCGCTCTAGAGGCGAACTTTTGT 3240  
DB 3181 AAAAAACAAATAGCCCTCTCCCTTCTTCCCTCCGCTCTAGAGGCGAACTTTTGT 3240  
QY 3241 TTCTACTCTTTTCAAGAGGGTTTTCTGTGTGTGTGGTTTTTGTCTTGTGTGACTG 3300  
DB 3241 TTCTACTCTTTTCAAGAGGGTTTTCTGTGTGTGTGGTTTTTGTGTGTGACTG 3300  
QY 3301 AAACAGAGAGTATTGACAGCAAAATCAGTAAACAAAGAAAGTAAATGCTTGGAGAG 3360  
DB 3301 AAACAGAGAGTATTGACAGCAAAATCAGTAAACAAAGAAAGTAAATGCTTGGAGAG 3360  
QY 3361 GAAAGGAGAGAGGAAAAATTTATATAAAACCTTAATAATTTGTTTTTTTTTCT 3420  
DB 3361 GAAAGGAGAGAGGAAAAATTTATATAAAACCTTAATAATTTGTTTTTTTTTCT 3420  
QY 3421 TTTCTATATATCTCTTGTGTGTGTCTAGCCTGATCAGATAGAGACAAACAGGAAG 3480  
DB 3421 TTTCTATATATCTCTTGTGTGTGTCTAGCCTGATCAGATAGAGACAAACAGGAAG 3480  
QY 3481 AATAGAGACCTCTGAGAGCAGAGTCTCTCCACCCCGAGAGTCTCAACAGCAC 3540  
DB 3481 AATAGAGACCTCTGAGAGCAGAGTCTCTCTCCACCCCGAGAGTCTCAACAGCAC 3540  
QY 3541 ATTCTGTGTCATCAAAACGAAACCAACTAGCAGAGGCGCTGAGAGAACACACACC 3600  
DB 3541 ATTCTGTGTCATCAAAACGAAACCAACTAGCAGAGGCGCTGAGAGAACACACACC 3600  
QY 3601 AGACACTTTTACAGTATTTTCAAGTGTCTTCCACACAGAAACCTTGAAAAAAACGATT 3660  
DB 3601 AGACACTTTTACAGTATTTTCAAGTGTCTTCCACACAGAAACCTTGAAAAAAACGATT 3660  
QY 3661 TCTAGAGCCGCTGTTAACCTCTGTTTACAGTTATATATATATATATATATATATAT 3720  
DB 3661 TCTAGAGCCGCTGTTAACCTCTGTTTACAGTTATATATATATATATATATATATAT 3720  
QY 3721 ATATATATAAAGGTACTGTAACTACTGTATACCTCCGACTTATATATATATATATAT 3780  
DB 3721 ATATATATAAAGGTACTGTAACTACTGTATACCTCCGACTTATATATATATATATAT 3780  
QY 3781 ACAGCAGATAGACAAAGACATAGCTTCGCGTCCGCTCTGTGTGCAAAAGGTTTACGCC 3840  
DB 3781 ACAGCAGATAGACAAAGACATAGCTTCGCGTCCGCTCTGTGTGCAAAAGGTTTACGCC 3840  
QY 3841 CAGGATGGGAGAGGGAGACCTGAGAGGGGTTTTAACAACTGAAGATGACCATAT 3900

Db 3841 CAGGATGGGAGAGGAGCAGCTGGAGGGGTTTAAACAACTGAAGATGACCCATAT 3900  
QY 3901 CACCCCCACCCCTGCCCCATGCTTACCTGCGCAAAAGGGGCTCAGCTGAGGTG 3960  
Db 3901 CACCCCCACCCCTGCCCCATGCTTACCTGCGCAAAAGGGGCTCAGCTGAGGTG 3960  
QY 3961 GTGGACCCGCGGAGCTGAGTGGGAATTTATCCAGACTCGCTGCAATAACCTTGA 4020  
Db 3961 GTGGACCCGCGGAGCTGAGTGGGAATTTATCCAGACTCGCTGCAATAACCTTGA 4020  
QY 4021 ATATGAATCTAAATGACTGCTCAGAAAATGCTTGAGAAAACATTTGCTCCTGATTT 4080  
Db 4021 ATATGAATCTAAATGACTGCTCAGAAAATGCTTGAGAAAACATTTGCTCCTGATTT 4080  
QY 4081 GAATTCGTGAGCACCTTGAAGGCCCTTGTGGGATCAGAAAATTTCCAGAGTGAGGAA 4140  
Db 4081 GAATTCGTGAGCACCTTGAAGGCCCTTGTGGGATCAGAAAATTTCCAGAGTGAGGAA 4140  
QY 4141 AGTGACCCGCGCATTAACCCCNCTGGAGCAATAAATAACATAACATAATGT 4192  
Db 4141 AGTGACCCGCGCATTAACCCCNCTGGAGCAATAAATAACATAACATAATGT 4192

RESULT 2  
AAZ37097  
ID AAZ37097 standard; DNA; 4416 BP.  
AC AAZ37097;  
XX  
DT 27-MAR-2000 (first entry)  
DE DNA encoding de novo DNA cytosine methyltransferase DNMT3A.  
KW De novo DNA cytosine methyltransferase; DNMT3A; neoplastic disorder;  
KW carcinoma; sarcoma; leukaemia; DNA methylation; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 353..3091  
FT /\*tag= a  
FT /product= "de novo DNA cytosine methyltransferase  
DNMT3A"

XX WO9967397-A1.  
XX  
XX 29-DEC-1999.  
XX  
XX 25-JUN-1999; 99WO-US14373.  
XX  
XX 25-JUN-1998; 98US-0090906.  
XX 24-JUL-1998; 98US-0093993.  
XX  
XX (GEHO ) GEN HOSPITAL CORP.  
XX  
XX Li E, Okano M, Xie S;  
XX  
XX WPI; 2000-106298/09.  
XX P-PSDB; AAY54057.  
XX  
XX New mouse and human polypeptides, useful to treat and diagnose  
PT neoplastic disorders e.g. carcinomas, sarcomas and leukemias -  
XX  
XX Claim 10; Fig 1C; 114pp; English.  
XX  
XX The present sequence encodes a human de novo DNA cytosine  
CC methyltransferase designated DNMT3A. The polypeptides can be administered  
CC therapeutically, especially by expressing encoding polynucleotides, to  
CC treat diseases associated with DNA cytosine methyltransferase, such as  
CC neoplastic disorders e.g. carcinomas, sarcomas and leukemias. They can  
CC be used to diagnose, or determine susceptibility to neoplastic disorders,  
CC by assaying for polypeptide expression levels in mammalian cells/body

CC fluids. They are useful to screen for compounds inhibiting/activating the  
CC polypeptide. The polypeptides can also be used for in vitro de novo  
CC methylation of DNA. Such in vitro methylation may be used to direct or  
CC regulate DNA expression in biological systems, e.g. recombinant DNA  
CC methylated in vitro may be introduced into a cell/organism to increase  
CC or decrease expression of a desired polypeptide for which the native DNA  
CC is under-methylated or not methylated. The polypeptides can also be  
CC used to produce antibodies which are useful to detect and purify the  
CC polypeptide or therapeutically e.g. to treat neoplastic disorders. The  
CC polynucleotides are useful to produce probes and primers which are  
CC useful diagnostically.  
XX  
SQ Sequence 4416 BP; 1107 A; 1131 C; 1327 G; 851 T; 0 other;

Query Match 68.5%; Score 2871.8; DB 21; Length 4416;  
Best Local Similarity 83.7%; Pred. No. 0;  
Matches 3574; Conservative 0; Mismatches 593; Indels 113; Gaps 24;  
QY 17 CCGGGCCGCGCCGAGCCGCGGCACACGCGAGAGCCGCTGAAAGCCAGCGTGAAGGT 76  
Db 147 CCGGGCCGCGCCGAGCCGCGGCATACGCTGGAGCCATCGAAGCCCCCACCACAGGT 206  
QY 77 G-----CACCTTTTCCGAGGGCTTGACATCAGGGTCTATGTTAAAGTCTTAGCTCT 126  
Db 207 GACAGAGGCACCGTTTACACAGAGGGCTCAACACCGGATCTATGTTAAAGTTTAACTCT 266  
QY 127 TGCTTACAAAGACACACGCAATTCCTTCTGAAGCCCTCGCAGCCCCACAGCGCCCTCG 186  
Db 267 CGCTCCAAAGACACCGATAATTCCTTCCCAAGCCC-AGCAGCCCCCCCAGC-CCCGCG 324  
QY 187 CAGCCCGCCGCTGCGCTTACTGCCCA-----GCAATGCCCTCCAGCGCCCGCGGA 239  
Db 325 CAGCCCGCCGCTGCTTCCGCGCCGAGATGCCGCGCATGCCCTCCAGCGCCCGCGGA 384  
QY 240 CACGAGAGCTCTCTCTGAGCGGGAGGATGATGAAAGGAGGAGAGGAGGAGGAGGAG 299  
Db 385 CACGAGAGCTCTCTGCGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 444  
QY 300 GAAACCGTGCAAGAGAGCGCCAGGAGCCAGCGCCACCGCCCGCGGAAGTGGGGAGGCC 359  
Db 445 GCGCGTGGCAAGGAGGAGCGCCAAAGAGCCAGACCCAGCGCAGGGAAGTGGGGCGGC 504  
QY 360 TGGCGGAGCGGAGCAGCACCACCGGTGGAAAGCAGTGCACACCCCGCAAGGAGCCAGCAGT 419  
Db 505 TGGGAGGAGCGCAAGCACCACCGGTGGAAAGCGTGCACACCGCCAAAGGAGCCCTCGGT 564  
QY 420 GACCAACAGTCTCAGCGCCATGCGCCAGGACTCTGGCCCTCAGATCTGCTACCAATGG 479  
Db 565 GATCTCCAAGTCCCCCATCCATGCCCCAGGACTCAGCGCTCAGAGCTATTACCAATGG 624  
QY 480 AGACTTTGAGAGGAGGAGTGAACCCCACTGAGGAGGGAGCCAGCTCAGGGCAGGA 539  
Db 625 GACTTTGAGAGGAGGAGTGAACCCCACTGAGGAGGGAGGAGGAGGAGGAGGAGGAG 684  
QY 540 GGGTGGGCGCCAGCTGAGAGAGGG---AACTGAGACCCCAACAGAGCTTCCAGAGC 596  
Db 685 GGGCGGGGCGCCAGCAGAGGGAGGGTGCAGTGCAGACCTTGCTGAGGCTTCAAGAGC 744  
QY 597 TGTGAGAAATGGCTGCTGTGTGACCAAGGAGCGCTGGAGCCTCTGACGAGAGGGCAA 656  
Db 745 AGTGAATAATGGCTGCTGCACCCCGCAAGGAGGGCGCGAGGAGCCCTGTCAGAGCGGCAA 804  
QY 657 AGAAACAGAGCAGACCAACATCGAATCCATGAAATGGAGGGGCTCCCGGGCGGAGTGG 716  
Db 805 AGAAACAGAGGAGACCAACATCGAATCCATGAAATGGAGGGCTCCCGGGCGGCTGGC 864  
QY 717 AGGTGGCTTGGGCTGGGAGTCCAGCTCCGTCAGCGAGCCCATGCGCAAGACTTCACCTTCCA 776  
Db 865 GGGTGGCTTGGGCTGGGAGTCCAGCTCCGTCAGCGGGCCCATGCGGAGGCTCACCTTCCA 924  
QY 777 GGCAGGGGACCCCTACTACATCAGCAAAACGGAAACGGATGAGTGGCTGGCAGCTTGGAA 836  
Db 925 GGGGGGGGACCCCTACTACTACATCAGCAAGCGCAAGCGGAGCGGATGGCTGGCACTGGAA 984

OY	837	AAAGAGGCTGAGAAAGAAAGCAAGCTAATTGGAAGTAATGATGCTGTGAAAGAGAACCA	896
Db	985	AAAGGAGGCTGAGAAAGAAAGCAAGGCTCATTTGAGAGAAATGAATGCTGTGAAAGAAACCA	1044
OY	897	GAGCCTCTGGAAGTCTCGAAGAGTGGAGAGGCGACGCCCTCTGCTGTGCAGACAGCCAC	956
Db	1045	GGGGGCCCGGGAGTCTCGAAGGTGGAGAGGCGACGCCCTCTGCTGTGCAGACAGCCAC	1104
OY	957	GAACCTGCTTCTCCGACTGTGGCAACACCCCTGAGCCAGTAGAGAGGGAGTGTGGAGA	1018
Db	1105	TGACCCCGCATCCCCCACTGTGGCTACACGCTCGAGCCCGTGGGGTCCGATGTGGGGA	1164
OY	1017	CAAGAAATGCTCCAAAGCAGCCGACGATGAGCTGAGTATGAGATGCTCCGGGCTTTGG	1076
Db	1165	CAAGAAATGCCCAAAACAGCGAGTGCAGCCAGAGTACAGAGACGCGCGGGGCTTTGG	1224
Db	1077	CATTGGAGAGCTGTGTGGGGGAAACTTGGGGCTTCTCTGGTGGCCAGGCCGAATGT	1136
Db	1225	CATTGGGAGGCTGTGTGGGGGAAACTGCGGGCTTCTCTGGTGGCCAGGCCGAATGT	1284
OY	1137	GTCCTTGTGATGTCAGGCGCGAGCCGAGCAGCTGAAAGCACTCGTGTGGTCAATGGTT	1186
Db	1285	GTCCTTGTGATGTCAGGCGCGAGCCGAGCAGCTGAAAGCACTCGTGTGGTCAATGGTT	1344
OY	1197	CGGAGATGGCAATTTCTCACTGTGTGTGTGAGAAAGCTCATGCGCTAGGCTCTTCTG	1256
Db	1345	CGGAGACGGCAAAATTTCTCACTGTGTGTGTGAGAAAGCTCATGCGCTAGGCTCTTCTG	1404
OY	1257	CAGGCAATTCACAGGCGCAACTACAAACAGACGCCCATATACCCGAAAGCCATCTACGA	1316
Db	1405	CAGGCGCTTCCACAGGCGCAACTATCAACAGACGCCCATATACCCGAAAGCCATCTACGA	1464
OY	1317	AGTCTCTCAGGTGGCCAGACAGCCGTGCGGGAGAGCTGTTTCAGCTTGCATGACAGTGA	1376
Db	1465	GGTCTCTGAGGTGGCCAGACAGCCGTGCGGGAGAGCTGTTTCAGCTTGCATGACAGTGA	1524
OY	1377	TGAAAGTGAACATGGCAAGGCTGTGAAAGTGCAAAACAAGCAGATGATTTGAAATGGAGCCCT	1436
Db	1525	TGAAGTGAACATGCTCAAGGCGCTGTGAGGTGCAGAAACAAGCCCATGATTTGAAATGGAGCCCT	1584
OY	1437	CGGTGGCTTCCAGCCCTCGGGTCTTAAAGGCCCTGGAGCCACCAAGAAAGAAACAATCC	1496
Db	1585	GGGGGGCTTCCAGCATTTATGAGCCCTAAAGGCCCTGGAGCCACCAAGAAAGAAACAATCC	1644
OY	1497	TTACAAAGGAAGTTTACACGCAATGTGGGTGGAGCCCTGAAGACAGCTGTACGCCCAAC	1556
Db	1645	CTACAAAGAGTGTATACGCAATGTGGGTGGAACTTGAAGCAGCTGTACGCCCAAC	1704
OY	1557	CCCAACAGCCCAAGAAACCCAGAAAGAGCAACAAGAGAAACTTAAGTTCAGAGATCAT	1616
Db	1705	TCCACCAAGCCCAAAAGGCCCGGAAAGAGCAACAGCGAGGAAAGCCCAAGTCAAGAGACATTAAT	1764
OY	1617	TGATGAGCGCAACAAGGAGCGGCTGGTGTATGAGGTGCCGACAGAGTGCAGAAACATGGA	1676
Db	1765	TGATGAGCGCAACAAGAGCGGCTGGTGTATGAGGTGCCGAGAAAGTGCAGGAAACATTTGA	1824
OY	1677	GGACATTTGTATCTCAATGTGGAGGCTCAATGACACCTGGAGAGCAACCACTTTTCAATGG	1736
Db	1825	GGACATTTGTATCTCTGTGGAGGCTCAATGATTTACCTTGGAAACCCCTTCTTGTGGTGG	1884
OY	1737	AGGCATGTGCAGAACTGTAAAGACTGCTTCTTGAGAGTGTGCTTACAGTATGACGACGA	1796
Db	1885	AGGATGTGCCAAAACTGCAAGAAATGCTTTCTTGAGAGTGTGCTTACAGTATGACGACGA	1944
OY	1797	TGGGTACCAATCTTATTTGCAACCATCTGTGTGGGGGGCGTGAAGTGTCTATGTGTGGAA	1856
Db	1945	CGGCTACCAAGTCTTATCTGCAACCATCTGTGTGGGGGGCGTGAAGTGTCTATGTGTGGAA	2004
OY	1857	CAACAACTGCGCAGGAGGCTTTTGTGTGCAAGTGTGGAGATCTTTGTGTGGGGCCAGAGGC	1916
Db	2005	CAACAACTGCTCAGGAGGCTTTTGTGTGAGAGTGTGTGACACTTTTGTGTGGGGCCAGGAGGC	2064

QY	1917	TGCTCAGGACCCATTAAAGAAAGCCCTTGAAACCTGCTACATGTCGGGCAATAAGGCAC	1976
Db	2065	TGCCAGGACCCATTAAAGAAAGCCCTTGAAACCTGCTACATGTCGGGCAATAAGGCATC	2124
QY	1977	CTATGAGCTGCTGGGAAGAGGAGAAAGCTGACCTTCTCCACTCCAGATGTTCTTTGGCCA	2038
Db	2125	CTTACGGGCTGCTGGGAGGGAAGGACCTGCCCCCTCCAGATGTTCTTGGCTTAA	2184
QY	2037	TAAACATGACCAAGGAATTTGACCCCTCCAAAGGTTTACCACCTGTGCCAGCTGAGAGAG	2096
Db	2185	TAAACCAAGACCAAGGAATTTGACCCCTCCAAAGGTTTACCACCTGTGCCAGCTGAGAAAG	2244
QY	2097	GAAACCCATCCGCGCTGCTCTCTCTTTGATGAGGAATGCTAAAGGCTCTGTGTCTGAA	2156
Db	2245	GAAACCCATCCGCGGTGCTGCTCTCTTTGATGAGGAATGCTAAAGGCTCTGTGTCTGAA	2304
QY	2157	GAACTGGGCACTCCAAATGACCGCTACATTTGCTCCAGAGGTGTGTAGAGATCCCATAC	2216
Db	2305	GAACTGGGCACTTCAGGTGAAACCGCTACATTTGCTCCAGAGGTGTGTAGAGATCCCATAC	2364
QY	2217	GATGGGCAATGATGCGGACCAAGGAAAGATCATGTACGTCGGGGACGTCGACAGCTCAC	2276
Db	2365	GATGGGCAATGATGCGGACCAAGGAAAGATCATGTACGTCGGGGACGTCGACAGCTCAC	2424
QY	2277	ACAGAAGCATATCCAGAGATGGGGCCCATTTCCACTGTGATTGAGAGCATCTCTGCA	2338
Db	2425	ACAGAAGCATATCCAGAGATGGGGCCCATTTCCACTGTGATTGAGGGGCATCTCTGCA	2484
QY	2337	TGACCTCTCAATTTGCAACCTCTGCGGCAAGGACCTTTATGAGGTTATCTGCGCGCTCTT	2386
Db	2485	TGACCTCTCAATCTCAACCTCTGCAAGGACCTTTACGAGGACATCTGCGCGCTCTT	2544
QY	2397	CTTTGAGTTCTACCGCCTCTCTGATATGCGGCGCCCAAGAGGAGATGATCGCCCTT	2456
Db	2545	CTTTGAGTTCTAACCGCCTCTCTGATATGCGGCGCCCAAGAGGAGATGATCGCCCTT	2604
QY	2457	CTTCTGGCTCTTTAGAAATGTGTGGCATGGGCGTTAGTACAAAGAGACATCTCGG	2516
Db	2605	CTTCTGGCTCTTTAGAAATGTGTGGCCATGGGCGTTAGTACAAAGAGGACATCTCGG	2654
QY	2517	ATTTCTTGAGCTTAAACCCCGTATATGATTAACGCCAAAGAAAGTGTCTGCTGCACACAGGC	2576
Db	2665	ATTTCTGAGCTCAACCCCTGTGATGATTAATGCAAGAAAGTGTGACCTGCACACAGGC	2724
QY	2577	CCGTTACTTCTGGGATTAACCTTCTGAGCAATGAACAGCCTTTGGGATCACTGTGAATGA	2636
Db	2725	CCGCTACTTCTGGGATTAACCTTCCCGATATGAACAGCCGTTTGGATCACTGTGAATGA	2784
QY	2637	TAACTGAGAGCTGCAGAGATGTCTGAGACACGGCAGAAATATGCCMATTTCAAGCAAAATGAG	2696
Db	2785	TAACTGAGAGCTGCAGAGATGTCTGAGACATGGCAGAGATATGCAAGATTCAGCAAAATGAG	2844
QY	2697	GACCATTAACCAACGAGTCAAACTCTATAAAGCAGGCAAAAGCCACGATTTCCCGCTTT	2756
Db	2845	GACCATTAACCAAGAGTCAAACTCCATAAAGCAGGCAAAAGCCACGCAATTTCCCTGCTTT	2904
QY	2757	CATGAACGAGAGGAGACATCTCTGTGTGCACTGAAATGAAAGAGGTGTGGCTCTCC	2816
Db	2905	CATGAATGAAGAAAGGACATCTTATGTGTCACTGAAATGAAAGAGATTTGGTTCTCC	2964
QY	2817	CGTCCACTACACAGACTCTTCCAAACATGACCGCTTTGGCGAAGCAGACATGCTGGGCG	2876
Db	2965	AGTCCACTATATCTGAGAGTCTTCCAAACATGACCGGCTTTGGGAGGCAAGACATGCTGGGCG	3024
QY	2877	ATCGTGAAGCGTGC CGGTCAATCCGCACTCTTCCGTCCGCTGAAGGAATATTTTGGCTTG	2936
Db	3025	GTCATGAGGCGTGCAGTCAATCCGCACTCTTCCGTCCGCTGAAGGAGTATTTTGGCGTG	3084
QY	2937	TGTGTAAAGGCAATGGGGCGAAACTGAATAGTGAATTAATAAAAGTTAAACAAACAAC	2996
Db	3085	TGTGTAAAGGCAATGGGGCGAAACTGAAGTACGA---CACAAAGTTTAAACAAACAAC	3140
QY	2997	AAACAAACAAACAAACAAATTAATAAACCAAGACGAGAGCGAGAAAAGT--TCA	3055



XX	Sequence 2938 BP; 759 A; 695 C; 774 G; 700 T; 10 other;	
80	Query Match 43.3%; Score 1814; DB 24; Length 2938; Best Local Similarity 82.5%; Pred. No. 0; Matches 2327; Conservative 5; Mismatches 398; Indels 89; Gaps 19;	
QY	1446 CCAAGCCCTGGGCTCTTAAGGGCCCTGAGCCCAACGAAAGAGAGAAATCTTTCATCAAGA	1505
DB	9 CCAAGCCCTTGGGCTCTTAAGGGCCCTGAGCCCAACGAAAGAGAGAAATCTTTCATCAAGA	68
QY	1506 AGTTTACACGACATGAGGGGAGGAGCCGTAAGCAGCTGTATAGCCCAACCCCAACGAGC	1565
DB	69 AGTTTACACGACATGAGGGGAGGAGCCGTAAGCAGCTGTATAGCCCAACCCCAACGAGC	128
QY	1566 CAAAGAAACCGAAGAAAGACAACAAGAGAAACCTTAAGGTCAAGAGATCAATTGATGACG	1625
DB	129 CAAAGAAACCGGGAAGAGACAAGCGAGAAAGCCAAAGGTCAAGAGATTAATTGATGACG	188
QY	1626 CACAGGAGCGGCTGGTGTATGAGGTGCGCAGAGGTGCAGAAACATCGAGACATTTG	1685
DB	189 CACAGAGAGACGGCTGGTGTATGAGGTGCGGACAGAGGTGCGGAAACATTTGAGACATCTG	248
QY	1686 TATCTCATGAGGAGCCTCAATGTCAACCTGAGACACCACTCTTCAATGGAAGCATGTG	1745
DB	249 CATCTCTGAGGAGCCTCAATGTACCTTGAACACCCCTCTTGTGAGAGAAATGTG	308
QY	1746 CCAAGACTGTAAGAACTGCTTCTTGAAGTGTCTTACCAATATGACAGCATGAGTACCA	1805
DB	309 CCAAAATCGCAAGAACTGCTTCTTGAAGTGTGCTTACCAATATGACAGCATGAGTACCA	368
QY	1806 GTCTTATGACCAATCTGCTGTGAGGGGCGTGAAGTCTCATGTTGTGAGAAACAACATG	1865
DB	369 GTCTTATGACCAATCTGCTGTGAGGGGCGGTAAGTCTCATGTTGTGAGAAACAACATG	428
QY	1866 CTGAGGTTGCTTTTGTGTCAAGTGTGTGATCTTGTGTGAGGCGCAGAGCTGCTCAAGC	1925
DB	429 CTGAGGTTGCTTTTGTGTCAAGTGTGTGATCTTGTGTGAGGCGGCGGCTGCTCAAGC	488
QY	1926 AGCCATTAAAGAAAGACCCCTGGAACCTCTCATGTGCGGACATMAAGGACCTTAAGGCT	1985
DB	489 AGCCATTAAAGAAAGACCCCTGGAACCTCTCATGTGTGCGGACCAAGGATCTTAAGGCT	548
QY	1986 GCTGCGAAGACGGGAGACTGCGCTTCTGCACTCCAGATGTTCTTTGCCAATAACCATGA	2045
DB	549 GCTGCGCGGCGAGAGAGACTGCGCTTCTGCACTCCAGATGTTCTTTGCCAATAACCATGA	608
QY	2046 CCAAGAAATTTGACCCCAAGTTTACCACTGTGCAAGTGAAGAGAGAGAGAGAGAGAG	2105
DB	609 CCAAGAAATTTGACCCCAAGTTTACCACTGTGCAAGTGAAGAGAGAGAGAGAGAGAG	668
QY	2106 CCGGATGCTGTCTCTTTGATGAGATTTGCTACAGGGCTCTGTGTGTAAGAGACCTGGG	2165
DB	669 CCGGATGCTGTCTCTTTGATGAGATTTGCTACAGGGCTCTGTGTGTAAGAGACCTGGG	728
QY	2166 CATCAAGTGAACCGCTACATTTGCTCGAGTGTGTGAGGACTCAATCAAGGTGGCAT	2225
DB	729 CATCAAGTGAACCGCTACATTTGCTCGAGTGTGTGAGGACTCAATCAAGGTGGCAT	788
QY	2226 GGTGCGGACCAAGGAAAGATCATGTACGTGCGGAGAGCTCGGACGCTCACAGAAACA	2285
DB	789 GGTGCGGACCAAGGAGAAATCATGTACGTGCGGAGAGCTCGGACGCTCACAGAAACA	848
QY	2286 TATCAGAGAGTGGGCGCCATTCGACGTGTGATTTGAGAGGAGCTCCCTGCAATGACCTTC	2345
DB	849 TATCAGAGAGTGGGCGCCATTCGATGTGTGATTTGAGAGGAGCTCCCTGCAATGACCTTC	908
QY	2346 CATGTCAACCTGCGCCGCAAGGACTTTATGAGGATATGCGCGCTCTTCTTTGAGTT	2405
DB	909 CATGTCAACCTGCTGCAAGGCGCTCTACGAGGACCTGCGGCGCTCTTCTTTGAGTT	968
QY	2406 CTACCGCTCTTGATATATGCGGCGCCCAAGAGAGAGATGATGCGCCCTTCTTTGAGCT	2465
DB	969 CTACCGCTCTCTCATGATGCGGCGCCCAAGAGAGAGATGATGCGCCCTTCTTTGAGCT	1028
QY	2466 CTTTGAAGATGTGTGAGCCCATGAGCGGTATGATGACAAGAGGACATCTCGGATTTCTGA	2525
DB	1029 CTTTGAAGATGTGTGAGCCCATGAGCGGTATGATGACAAGAGGACATCTCGGATTTCTGA	1088
QY	2526 GTCTTAACCCCGTATGATTTGACCCCAAGAGTGTCTGCTGCACACAGGCGCCGTTACTT	2585
DB	1089 GTCTTAACCCCGTATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG	1148
QY	2586 CTGGGGGTAACCTTCTGAGGACATGAGAGGCGCTTGGATCCACTGTAATGATTAAGCTGA	2645
DB	1149 CTGGGGGTAACCTTCTGAGGACATGAGAGGCGCTTGGATCCACTGTAATGATTAAGCTGA	1208
QY	2646 GCTGCAAGAGTGTGTGAGACCGGCAAGATAGCAAGTTTCAGCAAAAGTGAAGACCAATTAC	2705
DB	1209 GCTGCAAGAGTGTGTGAGACCGGCAAGATAGCAAGTTTCAGCAAAAGTGAAGACCAATTAC	1268
QY	2706 CACCAAGTCAAACTCTATTAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	2765
DB	1269 TACGAGGTCAAACTCTATTAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1328
QY	2766 GAAGAGGACATCTGTGTGCACTGAAATGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	2825
DB	1329 GAAGAGGACATCTGTGTGCACTGAAATGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1388
QY	2826 CACAGAGTCTCCAAATGAGACCGCTTGGCGAGGACAGAGACTGCTGGGCGGATCTGAGAG	2885
DB	1389 TACTGAGTCTCCAAATGAGACCGCTTGGCGAGGACAGAGACTGCTGGGCGGATCTGAGAG	1448
QY	2886 GGTGCGGCTCATCGCCACCTCTTCCGCTCGCTGAAGAAATATTTGCTGTGTGAAG	2945
DB	1449 GGTGCGGCTCATCGCCACCTCTTCCGCTCGCTGAAGAAATATTTGCTGTGTGAAG	1508
QY	2946 GACATGGGGGCAAACTGAAGTATGATGATTAATAAAAGTTTAAACAAACAAACAAACAA	3005
DB	1509 GACATGGGGGCAAACTGAAGTATGATGATTAATAAAAGTTTAAACAAACAAACAAACAA	1559
QY	3006 ACAAAACAAACAAATTAACCAAGAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	3064
DB	1560 ACAAAACAAACAAATTAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1619
QY	3065 AGAGAAAG	3122
DB	1620 AGAGAAAG	1679
QY	3123 TCGAAAGAGGTTGACATCTCTCTGAGTTTTCATATGTTAACTTCACTCATCTTAA	3182
DB	1680 TCGAAAGAGGTTGACATCTCTCTGAGTTTTCATATGTTAACTTCACTCATCTTAA	1739
QY	3183 AAGCAAAATAGAGC-CCCTCCCTTCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTT	3238
DB	1740 AAGCAAAATAGAGC-CCCTCCCTTCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTT	1799
QY	3239 TTTTCTACTTTTTCAGAGGGTTTCTGTTTGTGTTGTTTGTGTTTGTGTTTGTGTTTGTG	3298
DB	1800 TTTTCTACTT-TTTTAGAGGGTTTCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG	1857
QY	3299 TGAACCAAGAGATTTATGAGAGC-AAAATGATTAACAACAAAGTAAAGTAAAGTAAAGT	3357
DB	1858 TGAACCAAGAGATTTATGAGAGCAAAATGATTAACAACAAAGTAAAGTAAAGTAAAGT	1917
QY	3358 GAGAGAGG-----GAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	3408
DB	1918 GAGAGAGGAG	1977
QY	3409 TTTTCTTCTCTT-----TCTATATATCTCT	3435
DB	1978 TTTTCTTCTCTT-----TCTATATATCTCT	2037
QY	3436 TTGGTTGTCTTACCTGATCAATGAGAGACAAACA--GGAAGAGATAGAGACCTTC	3493
DB	2038 TTGGTTGTCTTACCTGATCAATGAGAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	2097



Db 1235 ACCTTCTCCAGCAGTCTCTGAGAGTCACTGGAGACCAAGCTGAAGCCCATCTGAGTGG 1294  
QY 1406 TGCAGAACAAAGCAGATGATTTGAATGGGCCCTCGGTGGCTTCCAGCCCTCGGCTCTAAG 1465  
Db 1295 GCCCAGCGTGGCTTCAAGCCTACTGAGGATTCGAGGGCTCAAAACCCACAAGAACCAACA 1354  
QY 1466 GCGTCGAGCCACCAAGAAAGAAAGAAATCTTTCAAGAAAGTTTACACCGACATGTGGG 1525  
Db 1355 GTGGTTATATGAGAAAGTGGCGCTGCTTCAAGACAGTAGGAATTTGAACCCAGAGAGCG 1414  
QY 1526 TGGAGCCTGAAG-----CAGCTGCTTACGCCCCACCCCAACCAAGAAAGCC 1575  
Db 1415 GAGAACAAAGTCCGAACACGACAACTATGCTGTGCTTGAAGTCCCCCAACC 1474  
QY 1576 AGAAAGACAACAAGAAACCTAAGTCAAGG---GATCATTGATGAGCGCACAAG 1632  
Db 1475 AAGCGCTCAAGACAAATAGTATGGGGAGAGACCGAGGGAGATGAGAGAGCCGA 1534  
QY 1633 GAGCGCTGGTGTATGAGGTGGCCAGAAAGTGCAGAAACATCGAGACATTTGTATCTCA 1692  
Db 1535 GAACGAGTGGCTTCTGAAGTCAACCAACAGGGCAATCGAAGACCGCTGTTGTCC 1594  
QY 1693 TGTGGAGCCTCAATGTCAACCCCTGGAGACCCACTCTTCAATTGGAGGCATGTGCCAAG 1752  
Db 1595 TGTGAAAAGAAACCTGTGTCTCTTCCACCCTCTTGAAGGTGGCTGTCAAGT 1654  
QY 1753 TGTAAAGATGCTTCTTGAAGTGTGTACCATGATGACGACGATGGGTACAGTCTCTAT 1812  
Db 1655 TGGCGGAGATGGCTTCTTAAGAGCTTCTACATGATATGATGAGAGCGGTATCAATCTTAC 1714  
QY 1813 TGCACCATCTGCTGTGGGGGGCGTGAAGTCTCATGTGTGGAAACAACCTGCTGACG 1872  
Db 1715 TGCACCTGTGTGTGAGGGCGGTGAACCTGCTGTGAGTACACAAAGCTGTGACGA 1774  
QY 1873 TGCCTTTGTGTCCAGGTGTGTGATCTCTGTGTGGGGCCAGAGTGTCTCAGAGCCATT 1932  
Db 1775 TGTCTTGTGTGAGTGTGTGAGAGTCTGTGTGGCCAGGACACAGCTGAGATGCCAAG 1834  
QY 1933 AAGGAAGACCCCTGGAAGTGTACATGTGCGGAGTAAAGGACCTTATGGGCTGTGGA 1992  
Db 1835 CTGCGAGAACCTGTGAGCTGTATATGTCTCTCTACAGCGCTGCCATGGGCTCTCCGA 1894  
QY 1993 AGACGGGAGACTGGCCTTCTGACCTCCAGATGTTCTTTTGGCAATTAACAATGACC--AG 2049  
Db 1895 CGCAGGAAAGTTGGAACATGCGGCTGCAAGACTTCTTCACTAATGATCCGACTGGAA 1954  
QY 2050 GAATTTACCCCCCAAGGTTTACCCACTGTGCCAGCTGAGAAAGAAAGCCATCCGC 2109  
Db 1955 GAAATTGAGCACCCAAATTGTACCCAGCAATCTCAGCCAAAGAGGCCCATTAAGA 2014  
QY 2110 GTGCTGTCTCTTTGATGGGATTGTACAGGGGCTCGGTGGTGGAGAGACCTGGGATC 2169  
Db 2015 GTCTGTCTCTGTTTGTATGGAATGCAACGGGATCACTTGGTGTCAAGAGATGGGATTT 2074  
QY 2170 CAAGTGAACCGCTACATTTGCTCCGAGTGTGTGAGACTCCATCAGGTGGCATGATG 2229  
Db 2075 AAGTGGAAAGTACATTTGCTCCGAAGTGTGTGACAGTGCATGCTGTGGGAACTGTT 2134  
QY 2230 CGGCACCAAGGAAAGATCATGTACGTGGGAGCTCCGACGCTGACACAGAACATATC 2289  
Db 2135 AAGCATTAAGCCAGATCAATATGTCAATGACCTCCGAAATAACCAAGAAATAATTT 2194  
QY 2290 CAGGAGTGGGGCCCATTCGACCTGTGATTTGGAGGAGTCCCTGCAATGACCTTCATT 2349  
Db 2195 GAAAGATGGGGCCCGTTGACCTGTGTATGTGTGAAAGCCCATCAATGATCTCTTAAC 2254  
QY 2350 GTCAACCTGTGCCGCAAGGAGCTTTATGAGGGTACTGCGCCTCTTCTTGAAGTTCTAC 2409  
Db 2255 GTCAATCTGTCCGCAAGGTTTATATGAGGCGACAGAAAGGCTCTTCTTGAGTTTAC 2314  
QY 2410 CGCTCTCTGATGATGCGCGGCCCAAGAGAGGAGATATGCGCCCTTCTTGTGCTCTTT 2469  
Db 2315 CACTTGTCTGAATATATACCCGCCCAAGAGGGCGACAACCGTCAATCTTCTGTGATGTT 2374

QY 2470 GAGATGTGTGGCCATGAGGCGTTAGTGAACAAGGAGACATTCGCGATTTCTTGAAGTCT 2529  
Db 2275 GAGATGTGTGGCCATGAAAGTGAATGACAAAGAAAGACTTCTCAAGATTCCTGGCATGT 2434  
QY 2530 AACCCGTGATGATTTGACGCCAAAGAAAGTGTCTGTGCACACAGGCGCGTATCTTGG 2589  
Db 2435 AACCCAGTATGATTCGATGTGCATCAAGTGTCTGTGCTCACAGGGCCCGGTACTTGTG 2494  
QY 2590 GGTAACTTCTGTGATGAACAGGCGTTTGGATCCACTGTGAAATGATAGCTGAGCTG 2649  
Db 2495 GTTAACCTACCCGAAATGAACAGGCGCGTATGAGCTTCAAGAAATGATAGCTGAGCTG 2554  
QY 2650 CAAGAGTGTGTGAGACAGGCAAGATAGCAAGTTCAGAAAGTGAAGGACCATTAACACC 2709  
Db 2555 CAGGACTGCTGTGAGTTCATATAGACGACAAAGTTAAAGAAAGTGCAGACAAATACACC 2614  
QY 2710 AGGTCAACTCTATTAAGCAGGCGCAAGAACCAAGATTTCCCGCTTTCATGAACGAGAAG 2769  
Db 2615 AAGTCGAATCTCATACAGACAGGCGCAAAACCAAGCTTTCCCTGTAGTCATGAATGCGAAG 2674  
QY 2770 GAGGACATCTGTGTGACACTGAATGGAAGGTTGTTGGCTTCCCGTCCACTACACA 2829  
Db 2675 GACGACGTTTGTGTGACACTGAGCTGAAAGATCTTGGCTTCCCTGCTCACTACACG 2734  
QY 2830 GACGCTCCAACTGAGAGCCGCTTGGCGAGGACAGACTGCTGGGCCGATGTGAGAGCGTG 2889  
Db 2735 GACGTGTCCAACTATGGGCGCGGCGCGCCGTCAGAAAGCTGTGGGACAGTCTTGAAGTGA 2794  
QY 2890 CCGGTATCCGCAACTCTTGTGCTCCGCTGAGAGAAATTTGCTTGTGTGTA 2942  
Db 2795 CCGGTATCAGACACGTTTGTGCTTGAAGAGACTACTTGTGCTGTGATTA 2847

RESULT 5  
AA237098  
ID AA237098 standard; DNA; 4145 BP.  
XX  
AC AA237098;  
XX  
XX 27-MAR-2000 (first entry)  
XX  
XX DNA encoding de novo DNA cytosine methyltransferase DNMT3B1.  
XX  
XX De novo DNA cytosine methyltransferase; DNMT3B1; neoplastic disorder;  
XX  
XX carcinoma; sarcoma; leukaemia; DNA methylation; ss.  
XX  
XX Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FT CDS 115..2676  
FT FT /tag= a  
FT FT /product= "de novo DNA cytosine methyltransferase  
DNMT3B1"  
FT exon 1..108  
FT FT /\*tag= b  
FT FT /number= 1  
FT exon 109..256  
FT FT /\*tag= c  
FT FT /number= 2  
FT FT 257..318  
FT FT /\*tag= d  
FT FT /number= 3  
FT FT 319..420  
FT exon 421..546  
FT FT /\*tag= e  
FT FT /number= 4  
FT exon 547..768  
FT FT /\*tag= f  
FT FT /number= 5  
FT exon 769..927  
FT exon

FT FT /tag= h  
FT /number= 7  
FT exon 928..1035  
FT /tag= i  
FT /number= 8  
FT exon 1036..1180  
FT /tag= j  
FT /number= 9  
FT exon 1181..1240  
FT /tag= k  
FT /number= 10  
FT exon 1241..1366  
FT /tag= l  
FT /number= 11  
FT exon 1367..1411  
FT /tag= m  
FT /number= 12  
FT exon 1412..1491  
FT /tag= n  
FT /number= 13  
FT exon 1492..1604  
FT /tag= o  
FT /number= 14  
FT exon 1605..1788  
FT /tag= p  
FT /number= 15  
FT exon 1789..1873  
FT /tag= q  
FT /number= 16  
FT exon 1874..2019  
FT /tag= r  
FT /number= 17  
FT exon 2020..2110  
FT /tag= s  
FT /number= 18  
FT exon 2111..2259  
FT /tag= t  
FT /number= 19  
FT exon 2260..2345  
FT /tag= u  
FT /number= 20  
FT exon 2346..2415  
FT /tag= v  
FT /number= 21  
FT exon 2416..2534  
FT /tag= w  
FT /number= 22  
FT exon 2535..4145  
FT /tag= x  
FT /number= 23  
XX  
XX WO967397-A1.  
XX  
XX 29-DEC-1999.  
XX  
XX 25-JUN-1999; 99WO-US14373.  
XX  
XX 25-JUN-1998; 98US-0090906.  
XX PR 24-JUL-1998; 98US-0093993.  
XX  
XX (GEO ) GEN HOSPITAL CORP.  
XX  
XX Li E, Okano M, Xie S;  
XX WPI; 2000-106298/09.  
XX DR P-PSDB; AAY54058.  
XX  
XX New mouse and human polypeptides, useful to treat and diagnose  
XX PT neoplastic disorders e.g. carcinomas, sarcomas and leukemias -  
XX PS Example 1; Fig 1D; 114pp; English.  
XX  
XX The present sequence encodes a human de novo DNA cytosine

CC methyltransferase designated DNMT3B1. The DNMT3B gene also produces,  
CC through alternate splicing, at least two shorter isoforms termed  
CC DNMT3B2 (comprising amino acids 1-355 and 376-853 of AAY54058, see  
CC AAY54059) and DNMT3B3 (sequences not given in the specification).  
CC of AAY54058, see AAY54060) (sequences not given in the specification).  
CC The polypeptides can be administered therapeutically, especially by  
CC expressing encoding polynucleotides, to treat diseases associated with  
CC DNA cytosine methyltransferase, such as neoplastic disorders e.g.  
CC carcinomas, sarcomas and leukemias. They can be used to diagnose, or  
CC determine susceptibility to neoplastic disorders, by assaying for  
CC polypeptide expression levels in mammalian cells/body fluids. They are  
CC useful to screen for compounds inhibiting/activating the polypeptide. The  
CC polypeptides can also be used for in vitro de novo methylation of DNA.  
CC Such in vitro methylation may be used to direct or regulate DNA  
CC expression in biological systems, e.g. recombinant DNA methylated in  
CC vitro may be introduced into a cell/organism to increase or decrease  
CC expression of a desired polypeptide for which the native DNA is  
CC under-methylated or not methylated. The polypeptides can also be used to  
CC produce antibodies which are useful to detect and purify the polypeptide  
CC or therapeutically e.g. to treat neoplastic disorders. The  
CC polynucleotides are useful to produce probes and primers which are useful  
CC diagnostically.  
XX  
XX SQ Sequence 4145 BP; 1041 A; 1083 C; 1096 G; 925 T; 0 other;  
Query Match 17.9%; Score 749.8; DB 21; Length 4145;  
Best Local Similarity 62.5%; Pred. No. 4.2e-172;  
Matches 1234; Conservative 0; Mismatches 717; Indels 24; Gaps 3;  
QY 992 AGCCAGTAGGAGGGGATGCTGGGGACAAAGATGCTACCAAGCAGCCGACGATGAGCCTG 1051  
DB 701 AGCAGGGGGCATGGAGTCCCGCAGGTGGAGGACAGTGGAGATGGAGACAGTTTCA 760  
QY 1052 AGTATGAGGATGGCGGGGCTTTGGCATTTGAGAGCTGCTGTGGGGGAACTTTGGGGCT 1111  
DB 761 AGTATCAGATGGGAGGAGTGGTGAATAGGGGACCTGCTGTGGGGAAAGATCAAGGGCT 820  
QY 1112 TCTCTGTGTGGCCAGGCCGAAATGCTCTTGTGTGATCAGAGCCGAGCCGAGCAGCTG 1171  
DB 821 TCTCTGTGTGGCCGCAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 880  
QY 1172 AAGGCACCTGCTGGGTCAATGTGTTCGAGATGGCAATTTCTCAGTGTGTGTGTGTGTG 1231  
DB 881 CTGGCATCGGTGGGTCCAGTGGTTTGGCGATGGCAAGTTTCTCGAGGTCTCTCGAGACA 940  
QY 1232 AGCTCATGCCCTGAGCTCCTTCTGCAAGTGCAATCCACAGGCCACCTACAAACAGCAGC 1291  
DB 941 AACTGTGGCCTGCGGCTGTTTTCAGCCAGCAGCTTTAATTTGGCCACCTTCAATAAGCTG 1000  
QY 1292 CCATGTACCGCAAGCCATCTACGAAAGTCTCTCAGGTGGCCAGCAGCCGTCGCGGAAG- 1350  
DB 1001 TCTCTATCGAAAGCCATGTACCATGCTCTGGAGAAAGCTAGGGTCTGAGCTGGAAGA 1060  
QY 1351 CTGTTTCCAGCTTGCATGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1397  
DB 1061 CTTTCCCGCAGCCCTGGAGACTCAATGGAGACCACTGAAGCCCATGTTGAGTGGG 1120  
QY 1398 ----TGTGGAAGTGCAGAACAGCAGATGATTTGAATGGGCCCTCGGTGGTTCAGCCCT 1453  
DB 1121 CCCACGGGGCTTCAGCCCACTGGGATCGAGGGCTCAAAACCAACCAACCAACCAACCAAC 1180  
QY 1454 CGGGTCTTAAGGGCTTGGAGCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1513  
DB 1181 TGGTTAATGAAGTCAAGGTGCTGTCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1240  
QY 1514 CCGACATGTGGGTGGAGCTTGAAGCAGCTGTTTACGCCCCACCCACCCACCCACCCACCCAC 1573  
DB 1241 AGAACCAAGACTCGAAGACGACAGCTGAGCAGCTGAGCAGCTGAGCAGCTGAGCAGCTG 1300  
QY 1574 CCAGAAAGAGCAGCAACAGAGAGAACTTAAGGTCAAGGAGATC-----ATTGATGAGCGCA 1627  
DB 1301 CCAAGCGCTCAAGACAAATTTGCTATAACACAGCGCAAGACCGAGGGGATGAAGATCAGA 1360



QY 1628 CAAGGAGCGGCTGTGTATGATGAGTGGCCAGAGTGCAGAAACATGAGGACATTGTGA 1687  
 DB 1361 GCCGAGAAACAATGCTTCAGATGTGCCAACAAAGAGAGCTGGAAGATGCGTGT 1420  
 QY 1688 TCTCATGTGGAGAGCTCATATGTCACTGGAGCACCCACTTTCATTGGAGGCACTGTGCC 1747  
 DB 1421 TGTCTTGTGGCAGGAAAAAACCCTGTCTCCACCTCTCTTTGAGGGGGGCTCTGTTC 1480  
 QY 1748 AGAATGTGAAGACTGCTTTCTTGGAGTGTGCTTACAGATGACAGACATGGGTACCACT 1807  
 DB 1481 AGACATCCGGGATCGCTTCTTCTTGTAGCTGTTTACATGTATGATGACATGCGTACAGT 1540  
 QY 1808 CCAATTCACCATCTGCTGTGTGGGGGGGTGAAGTCTCACTGTGTGGAAACAACACTGCT 1867  
 DB 1541 CTTACTGCACTGTGTGTGTGGAGGGCCGAGAGCTGTCTTGTGAGCAACACGACTGTCT 1600  
 QY 1868 GCAGGTCTTTTGTGTGAGTGTGATGCTCTTGTGTGGGGCCAGAGCTGTCAAGCAG 1927  
 DB 1601 GCCGCTGTCTGTGTGAGTGTCTGAGAGTGTCTGTGTGGGACAGGACAGCGGCCAGAG 1660  
 QY 1928 CCAATTGAAGAAAGCCCTGGAACTGTACATGTGCGGGCAATTAAGGCACTATGGCTGC 1987  
 DB 1661 CCAAGCTTCAGAGAGCTGTGAGCTGTACATGTGTCTCCGACAGCTGTCTATGCGCTCC 1720  
 QY 1988 TGGGAAGACGGGAAGACTGGCTTCTTCTGACTCCAGATGTTCTTTGCCAATTAACCATACC 2047  
 DB 1721 TGGGGCCCGCGAAGAGACTGTGAACTGTGCGCTGTGAGGCTTCTTACACAGTGAACAGGGGC 1780  
 QY 2048 AGGAATTGACCCCGCAAGGTTTACCCACCTGTGCCAGCTGTGAAGAAAGAAAGGCCCATCC 2107  
 DB 1781 TTGAAATGCAAGAGCCCGCAAGCTGTACCTGTGCAATTCGCCGAGCCGCAAGGGGCCATTC 1840  
 QY 2108 GCGTGTGTCTCTCTTGTATGAGATGCTCAAGGGCTCTGTGTCTGAAGGACCTGGGCA 2167  
 DB 1841 GAGTCTGTATGTGTGATGAGCATGCGACAGGCTACCTAGTCTCAAGAGATGTGGCA 1960  
 QY 2168 TCCAGTGAACCGGTACATGCTCTCGAGGTGTGTGAGAGCTCATGACGGTGGGCAATGG 2227  
 DB 1901 TAAAGGTAGGAAGATGACGTGCTTCTGAAGTGTGTGAGAGTCCATGTGTGGAACCG 1960  
 QY 2228 TGGGGCACCGAGGAAGATATGATGCTCGGGAACCTCCGAGGTCAACACAGAGATA 2287  
 DB 1961 TGAAGCACGAGGGGAATATCAAAATGATGAACGAGTGAAGATCAACAAAGAAATA 2020  
 QY 2288 TCCAGAGTGGGGGCCATTGCACTGTGTGATGTGAGGCAAGCTCTGCAATGACCTTCCA 2347  
 DB 2021 TTGAAAGATGGGGCCATTGACTGTGTGATGTGCGAAGGCCCATGCAACGATTTCTCAA 2080  
 QY 2348 TTGTCAACCTGTCCCGCAGAGGACTTATGAGGCTACGTGGCCGCTCTTTTGAATTC 2407  
 DB 2081 ATGTGAATCCAGCCAGAAAGGCTGTATGAGGTACAGGCCGCTCTTCTTGAATTTT 2140  
 QY 2408 ACCGCTCTCTGATGATGCGCGGCCCAAGAGGAGATGATCGCCCTTCTTGTGCTCT 2467  
 DB 2141 ACCACCTGTGAATTAATCTCAACGCCCAAGAGGGGTGTGACCGGCCCTTCTTGTGATGT 2200  
 QY 2468 TTGAGATGTGTGAGGAGGAGGTGTGTGACAGAGAGGACATGCGCATTTCTTGTAGT 2527  
 DB 2201 TTGAGATGTGTGAGGAGGAGGTGTGTGACAGAGAGGACATGCGCATTTCTTGTAGT 2280  
 QY 2528 CTAAACCCGATGATGATGACGCCAAGAGATGTGTGTGACACAGAGGCCGCTTACTTCT 2587  
 DB 2261 GTATATCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2320  
 QY 2588 GGGGTAACCTTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2647  
 DB 2321 GGGGTAACCTTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2380  
 QY 2648 TGCAGAGTGTGTGAGAGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2707  
 DB 2381 TGCAGAGTGTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2440  
 QY 2708 CCAAGTCAAACTTATTAAGCAGGCAAGACAGCAATTTCCCGTCTTCAATGAACGAGA 2767

DB 2441 CCAAGTCAAACTTATTAAGCAGGCAAGACAGCAATTTCCCGTCTTCAATGAATGCA 2500  
 QY 2768 AGGAGACATCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2827  
 DB 2501 MAGAAGATGTTTGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2560  
 QY 2828 CAGACGTCTTCAACATGAGCGGCTTGGCAGAGCAGACATGCTGGGCCGATGTGAGCGC 2887  
 DB 2561 CAGACGTCTTCAACATGAGCGGCTTGGCAGAGCAGACATGCTGGGCCGATGTGAGCGC 2620  
 QY 2888 TGGCGGTCTATCCGCGCACCTTGTGCTTCCGCTGAAGAAATATTTTGTGTGTGA 2942  
 DB 2621 TGCCTGTATCCGACACCTCTGCGCCCTTGTGAAGAGCTACTTGTGATGTGAATA 2675

## RESULT 6

AB057540  
 ID AB057540 standard; cDNA; 709 BP.

XX AB057540;

DT 02-AUG-2002 (first entry)

DE Human colon cancer related nucleotide sequence SEQ ID NO:1235.

KM Human; colon cancer; cancer; tissue profiling; forensic; mapping;

KM genetic analysis; diagnostic; antisense therapy; gene; ss.

OS Homo sapiens.

PN W0200229086-A2.

PD 11-APR-2002.

PF 02-OCT-2001; 2001WO-US30732.

PR 02-OCT-2000; 2000US-237271P.

PA (FARB ) BAYER CORP.

PI Burgess C, Aetle JH, Carroll E, Catino TJ, Dwiwedi P, Molino GA;

PI Thielingam A, Lewis ME;

XX WPI; 2002-426115/45.

PT New isolated nucleic acid that is differentially expressed in cancer  
 PT tissues useful for determining the presence of colon cancer in a cell  
 PT or tissue type, and in antisense therapy -

PS Claim 1; Fig 1; 796pp; English.

CC AB05306 to AB060787 represent isolated nucleic acids (I) differentially  
 CC expressed in cancer tissues. AB078993 to AB079004 represent proteins  
 CC encoded by the AB060776 to AB060787 nucleic acid sequences. (I) can be  
 CC used in antisense therapy. An antibody immunoreactive with a polypeptide  
 CC encoded by (I) is useful for detecting cancer in a patient sample, and  
 CC for detecting the presence or absence of a polynucleotide encoded by a  
 CC nucleic acid which hybridizes to (I) in a cell. A probe/primer derived  
 CC from (I) can be used for determining the presence of a nucleic acid which  
 CC hybridizes to (I), and for determining the phenotype of cells in a sample  
 CC of cells from a patient. (I) is useful for determining the presence of  
 CC colon cancer in a cell or tissue type, for determining the presence or  
 CC state of other type of cancer, in antisense therapy, to generate  
 CC macroarrays on a solid surface, to identify a chromosome on which the  
 CC corresponding gene resides, and in tissue profiling, forensics, genetic  
 CC analysis, mapping and diagnostic applications. (I) can be used to raise  
 CC antibodies, and to screen for peptide analogues and antagonists.

SQ Sequence 709 BP; 163 A; 177 C; 193 G; 168 T; 8 other;

Query Match 13.3%; Score 559; DB 24; Length 709;  
 Best Local Similarity 88.8%; Pred. No. 5.5e-126;

Matches 624; Conservative 0; Mismatches 73; Indels 6; Gaps 2;	
Qy 2250	GTACGTCGGGGACGTCCTCCAGCGTACACAGAGCATATCCAGAGTGGGGCCATTCGA 2309
Db 1	GGACGTCGGGGACGTCCTCCAGCGTACACAGAGCATATCCAGAGTGGGGCCATTCGA 60
Qy 2310	CTGGTGAATGGAGGAGTCCTCGAATGACCTCTCATGTCAACCTGCGCCAGGG 2369
Db 61	TCGGTGAATGGGGGAGTCCTCGAATGACCTCTCATGTCAACCTGCTCCAGGG 120
Qy 2370	ACTTTATGAGGTAATGCGCGCTCTCTTTGAGTTCTAGCCGCTTCCTGATGATGGCG 2429
Db 121	CCCTCTAGAGGGCACTGGCGCGCTCTCTTTGAGTTCTAGCCGCTTCCTGATGATGGCG 180
Qy 2430	GCCCAAGGAGGAGATGATCGCCCTCTCTGCTCTTGAAGATGGTGGCCATGGG 2489
Db 181	GCCCAAGGAGGAGATGATCGCCCTCTCTGCTCTTGAAGATGGTGGCCATGGG 240
Qy 2490	CGTTAGTGACAGAGGACATCGCGATTTCTTGAGTCTTAACCCCGTGATGATGACG 2549
Db 241	CGTTAGTGACAGAGGACATCGCGATTTCTTGAGTCTTAACCCCGTGATGATGATG 300
Qy 2550	CAAGAAGTCTCTGTCACAGCGCCGCTTACTTCTGGGTAACTTCTGTCATGAA 2609
Db 301	CAAGAAGTCTCTGTCACAGCGCCGCTTACTTCTGGGTAACTTCTGTCATGAA 360
Qy 2610	CAGCCCTTGGATCACTGAATGAAGCTGGAGCTGCAAGTGTCTGGAGCAGG 2669
Db 361	CAGCCCTTGGATCACTGAATGAAGCTGGAGCTGCAAGTGTCTGGAGCAGG 420
Qy 2670	CAGATAGCAAGCTTCAGAAAGTCAGGACCATTTACCACAGCTCAAACTCTATAAGCA 2729
Db 421	CAGATAGCAAGCTTCAGCAAGTCAGGACCATTTACTACGAGCTCAAACTCTATAAGCA 480
Qy 2730	GGGCAAGACAGCATTTCCCGCTCTCATGACGAGGAGGACATCTCTGTGT-GCA 2788
Db 481	GGGCAAGACAGCATTTCCCTGCTCTCATGAATGAAAAAGAGGACATCTTATGTGGCA 540
Qy 2789	CTGAATGGAAAGGTTGTGGCTTCCCGCTCCACTACAGACGCTCTCCAACATGAGCC 2848
Db 541	CTGAATGGAAAGGTTATTTGGTTCCCACTCCACTATCTACGCTCTTCAACATGANCC 600
Qy 2849	CGTTGGCGAGCAGAGACTGCTGGCGCGATCGT-----GGAGGCTGCCGTATCCCGCA 2903
Db 601	CGTTGGCGAGCAAAACTGCTGGCGCGTCAATTGGACCGTGCCAAATTCATCCNCACT 660
Qy 2904	CTCTCTCGCTCCGCTGAAGGAATATTTTCTGTGTGTAAGG 2946
Db 661	TNNTTNGNTTCCGTTGAAGGAATATTTTCTGTGTGTTAAG 703
RESULT 7	
AAH33860	
ID	AAH33860 standard; cDNA; 1060 BP.
XX	
AC	AAH33860;
XX	
DT	03-SEP-2001 (first entry)
XX	
DE	Human colon cancer antigen encoding cDNA SEQ ID NO:916.
XX	
KW	Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW	colorectal carcinoma; sa.
XX	
OS	Homo sapiens.
XX	
PN	WO200122920-A2.
XX	
PD	05-APR-2001.
XX	
PF	28-SEP-2000; 2000WO-US26524.
XX	
PR	29-SEP-1999; 99US-0157137.

PR XX

03-NOV-1999; 99US-0163280.

(HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Barash SC, Birse CE, Rosen CA;

WPI; 2001-235357/24.

P-PSDB; AAG74429.

Nucleic acids encoding 4277 human colon cancer-associated polypeptides, useful for preventing, diagnosing and/or treating colorectal cancers -

Claim 1; Page 2829; 9803pp; English.

AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patients own production of P. Additionally, N may be used to produce the colon cancer-associated Ps by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAH77789 represent sequences used in the exemplification of the present invention.

N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX SQ

Sequence 1060 BP; 267 A; 262 C; 297 G; 227 T; 7 other;

Query Match 13.1%; Score 547.6; DB 22; Length 1060;

Best Local Similarity 72.3%; Pred. No. 4.1e-123;

Matches 703; Conservative 4; Mismatches 266; Indels 0; Gaps 0;

Qy 1970

AGGGACCTATGGCTCTCGAAGACGGGAAGACTGGCCCTTCGACTCCAGATTTCT 2029

Db 9

AGCGCTGTCATGGCGTCTCGGCGCCGGAAGACTGGAAGCTGGCTCGAGCCCTTC 68

Qy 2030

TTGCCAATAACCATGACACAGGAATTTTGACCCGCCAAAGTTTACCACTGTGCGAGCTG 2089

Db 69

TCACCAGTGACACGGGGCTTGAATACGAAGCCCCCAAGCTGTACCTGCCATTCGCGAG 128

Qy 2090

AGAAGAGGAAGCCCATCCGGTGCTGTCTTTTGTGAGGATTTCTACAGGCTCTGG 2149

Db 129

CCGAAAGGCGGCCCATTCGAGTCTGTCTATTGTTGTGTGGCATTCGACAGGCTACTAG 188

Qy 2150

TGCTGAAGGACCTGGCATCCAGTGACCGCTACATTTGCTCCGAGGTGTGTGAGGACT 2209

Db 189

TCCTCAAGGTGTGGGCATAAAGTAGGAAGTACCTCGCTTCTGAAGTGTGTGAGGAGT 248

Qy 2210

CCATCAAGTGGGCTGTGGCGACGAGGAAGATCATGTAGTCGGGAGCTCCGCA 2269

Db 249

CCATTCTGTGTGAACCGTGAAGCAGGAGGATATCAATACGTGAACGATGAGGA 308

Qy 2270

GCGTCAAGAGCATATCCAGAGTGGGCGCCATTGCACTGTGTGATTTGGAGGAGTC 2329

Db 309

ACATCAAGAAATAATTCAGATGGGGCCCATTTGACTTGGTGTGTTGGCGAAGCC 368

Qy 2330

CCTGCAATGACCTCTCAATTTGAACCTGCGCAGGAGGACTTTATGAGGCTACTGACC 2389

Db 369

CATGCAAGCTCTCTCAATTTGAATCCAGCAGGAAGGCTGTATGAGGTTACAGGCC 428

Qy 2390

GCTCTCTTTTGAGTTCTTACCGCTCTGATGTCGCGGCCCAAGGAGGAGATGATC 2449

Db 429

GGCTCTTCTGAAATTTTACCACTGCTGAATTTACTCACGCCCCCAAGGAGGTTGATGACC 488

Qy 2450

GCCCCCTCTCTGCTCTTTTGAGAAATGTGTGGCCATGTGTAGTGTGCAAGAGGACA 2509





PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246529.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PA Rosen CA, Barash SC, Ruben SM;  
DR WPI; 2001-46570/50.  
DR P-PSDB; AAM94301.  
XX  
PT Isolated nucleic acid molecule encoding a reproductive system antigen  
XX is used in preventing, treating or ameliorating a medical condition -  
XX  
XX  
XX Claim 1; SEQ ID NO 272; 1297bp + Sequence Listing; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a coding sequence of the  
CC invention.  
XX  
XX Sequence 622 BP; 154 A; 161 C; 189 G; 112 T; 6 other;  
SQ  
Query Match 9.9%; Score 413; DB 22; Length 622;  
Best Local Similarity 87.7%; Pred. No. 2.1e-90;  
Matches 471; Conservative 2; Mismatches 62; Indels 2; Gaps 2;  
OY 1446 CCAAGCCCTCGGATCTTAAGGCGCTGAGCCACAGAGAAGAGAATCCTTACAAGGA 1505  
DB 2 CCAAGCCCTTGGGCTTAAGGCGCTGAGCCACAGAGAAGAGAATCCTTACAAGGA 61  
OY 1506 AGTTACACCGACATGTGGTGAGCCTGAAGACAGCTGCTTACGCCCCACCCACCGAGC 1565

DB 62 AGTGTACACCGACATGTGGTGAGCCTGAGCGCGCTGCTTACCGACACCTTACACCGAGC 121  
OY 1566 CCAAGAAACCCAGAAAGAGCAACAAGAGAAACCTTAAGGTCAAGGATTCATTGTAGCGC 1625  
DB 122 CAAAAGCCCCGGAGAGGACACAGCGAGAGAGCCCAAGGTCAAGAGATTAATTGTAGCGC 181  
OY 1626 CACAAGGAGCGGCTGTGTATGATGAGTGCGCCAGAAAGTGCAGAAACATCGAGACATTTG 1685  
DB 182 CACAAGAGAGCGGCTGTGTATGATGAGTGCGCGAGAAATGCGGAACTTAGAGCACTTG 241  
OY 1686 TATTCATGTGGAGCCTCATATGTCACCTTGAAGACACCACTTTCAATTGAGGCATGTG 1745  
DB 242 CATCTCCTGTGGAGCCTCAATGTTACCTTGAACACCCCTTTCCTTGGAGGAATGTG 301  
OY 1746 CCAAGACTGAAGACGCTTCTTGAAGTGCTTACCAATGAGACGATGGGTACA 1805  
DB 302 CCAAACTGCAAGACTGCTTCTTGAAGTGCTTACCAATGAGACGATGGGTACA 361  
OY 1806 GTCCATTGACACCATCTGCTGTGGGGGCGTGAAGTGCTCATGTGTGGAAACAACACTG 1865  
DB 362 GTCCACTGACACCATCTGCTGTGGGGGCGCTGAAGTGCTCATGTGTGGAAACAACACTG 421  
OY 1866 CTGAGGTGCTTTTGTGTGCTGAGTGATGATCTTGTGTGGGGCCAGAGCTGCTAGGC 1925  
DB 422 CTGAGGTGCTTTTGTGTGCTGAGTGATGATCTTGTGTGGGGCCAGAGCTGCTAGGC 480  
OY 1926 AGGCATTGAAGAAAGACCCCTGGAACGCTCAATGTGGGGCATTAAGGCACTTATG 1982  
DB 481 AGSCATTGAAGAAAGACCCCTGGAACGCTCAATGTGCGGACAAAGGATCTTACGG 536

RESULT 10  
AAS86165  
ID AAS86165 standard; cDNA; 711 BP.  
XX  
XX AAS86165;  
AC  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #21969.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200175067-A2.  
PN  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001MO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX  
XX WPI; 2001-639362/73.  
DR P-PSDB; ABG21978.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT reproducible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX Claim 1; SEQ ID NO 21969; 103bp; English.  
PS  
XX The invention relates to isolated polynucleotide (i) and  
CC polypeptide (ii) sequences. (i) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS84197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 711 BP; 181 A; 164 C; 199 G; 167 T; 0 other;

Query Match 8.1%; Score 341; DB 23; Length 711;  
Best Local Similarity 74.7%; Pred. No. 7.9e-73;  
Matches 428; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 2110 GTGCTGCTCTCTTTGATGGGATGCTACAGGGCTCTGCTGCTGAAGGACCTGGGCATC 2169

DB 1 GTCTGCTCATTTGTTGATGGCATCGCACAGGCTACTAGTCTCAAAGAGTTGGGCATA 60

QY 2170 CAAGTGCACCGCTACATTCCTCCGAGTGTGTGAGGACTCCATCAGGTGGGATGGTG 2229

DB 61 AAGTAGGAAGTACGTGCTCTGTAAGTGTGTGAGGAGTCCATTGCTTTGGAACCGTG 120

QY 2230 CGGCACACGGGAAGATCATGTAGTCTGGGAGCTCGGAGGCTCACAGACATATC 2289

DB 121 AAGCACAGGGGATATCAATCATCTGACGCTGAGGACATCAACAAGAAATATTT 180

QY 2290 CAGAGTGGGGCCCATTCGACTGTGTGATGTGGAGGAGTCCCTGCAATGACTCTCCATT 2349

DB 181 GAAGAATGGGGCCCATTTGACTGTGTGATGTGGGGAAGCCATGCAACGATCTTCAAT 240

QY 2350 GTCAACCTTCGCCGCAAGGACTTTATGAGGGTACTGGCGGCTCTCTTTGAGTTCTAC 2409

DB 241 GTGAATCCAGCCAGGAAAGGCTGTATGAGGGTACAGGCGGCTCTTCTCGAATTTTAC 300

QY 2410 CGCTCTCTGATGATGCGCGGCCCAAGGAGGAGATGATCGCCCTTCTTCTGGCTCTTT 2469

DB 301 CACCTGCTGAATTACTACCGCCCAAGGAGGGTATGACCGGCGCTTCTTCTGGATGTTT 360

QY 2470 GAGAATGTGTGGCCATGGCGTTAGTGACAAGAGGACATCTCGCGATTTCTTGAGTCT 2529

DB 361 GAGAATGTTGTAGCCATGAAGTGTGGCGACAAGAGGAGACATCTCACGGTTCTTGAGTGT 420

QY 2530 AACCCGCTGATGATGACGCCAAGAAGTGTCTGCTGCACACAGGCGCGTTACTTCTGG 2589

DB 421 AATCCAGTGNATGATGATGCCATCAAGTCTTCTGCTCTCACAGGCGCGATCTTCTGG 480

QY 2590 GGTAACCTTCCTGGCATGACAGGCTTTGGCATCCACTGTGATGATGAAGCTGAGCTG 2649

DB 481 GGCAACCTACCCGGGATGAACAGGCCCGTGATAGCATCAAGAATGATTAATCTGAGCTG 540

QY 2650 CAAGAGTGTCTGAGCAGCGCAGATAGCCCAAG 2682

DB 541 CAGGACTGCTTGGAATACAAATAGGATAGCCCAAG 573

RESULT 11

AAF80537

ID AAF80537 standard; cDNA; 2077 BP.

XX AAF80537;

XX AAF80537;

DT 08-JUN-2001 (first entry)

XX

DE Receptor #25 partial coding sequence.

XX Probe; microarray; cancer; immunopathology; neuropathology; ss.

XX Balaena mysticetus.

XX US6183968-B1.

XX 06-FEB-2001.

XX 25-MAR-1999; 99US-0276531.

XX 27-MAR-1998; 98US-0079677.

XX (INCY-) INCYTE PHARM INC.

XX Bandman O, Lal P, Hillman JL, Yue H, Reddy R, Guegler KJ;

XX Baughn MR;

XX WPI; 2001-201999/20.

XX Composition having probes which comprise part of gene sequence encoding  
XX proteins associated with cell proliferation useful as hybridizable  
XX array elements in Microarrays to monitor expression of target  
XX polynucleotide

XX Claim 1; Columns 93-96; 104pp; English.

XX The present invention relates to a composition comprising several  
XX polynucleotide probes. Probes can be derived from the present sequence.  
XX The probes are immobilized and are preferably useful as hybridizable  
XX array elements in a microarray for monitoring the expression of several  
XX polynucleotides. The microarray can be used in the diagnosis of cancer  
XX such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma  
XX and teratocarcinoma etc., immunopathology such as AIDS, Addison's  
XX disease, adult respiratory distress syndrome, allergies, anasmi, asthma,  
XX atherosclerosis and bronchitis etc., neuropathology such as Alzheimer's  
XX disease, amnesia, amyotrophic lateral sclerosis, bipolar disorder,  
XX catatonia and cerebral neoplasms etc. The microarray can also be used to  
XX investigate an individual's predisposition to a disease such as cancer,  
XX immunopathology or neuropathology. Also, the microarray can be used for  
XX investigating cellular response to infection, drug treatment etc. The  
XX microarray can be used for diagnostics, prognostics and treatment  
XX regimens, drug discovery and development, toxicological and  
XX carcinogenicity studies, forensics, pharmacogenomics etc. The array can  
XX also be used for monitoring disease progression.

XX Sequence 2077 BP; 539 A; 485 C; 477 G; 567 T; 9 other;

Query Match 7.7%; Score 324.6; DB 22; Length 2077;

Best Local Similarity 74.8%; Pred. No. 1.4e-68;

Matches 416; Conservative 0; Mismatches 139; Indels 1; Gaps 1;

QY 2387 GCGGCTCTCTTTGAGTCTTACCGCTCTCTCATGATCGCGCCCAAGGAGGATG 2446

DB 62 GCGGCTCTCTTTGAGTCTTACCGCTCTCTCATGATCGCGCCCAAGGAGGATG 120

QY 2447 ATGCGGCTCTCTTTGAGTCTTACCGCTCTCTCATGATCGCGCCCAAGGAGG 2506

DB 121 ACGGCGCTCTCTTTGAGTCTTACCGCTCTCTCATGATCGCGCCCAAGGAGG 180

QY 2507 ACATCTCGGATCTTCTGAGTCTTACCGCTCTCTCATGATCGCGCCCAAGGAGG 2566

DB 181 ACATCTCGGATCTTCTGAGTCTTACCGCTCTCTCATGATCGCGCCCAAGGAGG 240

QY 2567 CACACAGGCGGCTTACTTCTGGGTAACCTTCTGGCATGAACAGGCGCTTGGCATCA 2626

DB 241 CTCACAGGCGGCTTACTTCTGGGTAACCTTCTGGCATGAACAGGCGCTTGGCATCA 300

QY 2627 CTGTGAATGATAAGCTGGAGCTCAAGAGTGTCTGGAGCAGCGGAGAGTACCAAGTTCA 2686

DB 301 CAAAGAGTATAAATCTCGGCTGCAGGACTGTCTTGGAAATACAATAGGATAGCAAGTTAA 360

QY 2687 GCAAGTGGAGACCATTTACCAAGTCAACTCTATTAAGCAGGSCAAAGACGACATT 2746  
DB 361 AGAAAGTACAGACATTAACCAACCAAGTCAACTCTATTAAGCAGGSCAAAGACGACATT 420  
QY 2747 TCCCGCTTCAAGACAGAGAGAGAGACATCCCTGTGTGTCACTGAAGTGAAGGCTGT 2806  
DB 421 TCCCTGTGTGTCAAG 480  
QY 2807 TTGGCTTCCCGCTTCAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2866  
DB 481 TTGGCTTCCCGCTTCAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
QY 2867 TGGTGGGCGCATGTGGAGGCGTGGCTGATCCGCACTTTGCTCGCTGAAGGAT 2926  
DB 541 TGGTGGGAGAGTCTGGAGGCGTGGCTGATCCGCACTTTGCTCGCTGAAGGACT 600  
2927 ATTTCCTGTGTGTGA 2942  
601 ACTTGTGATGTGAATA 616  
b1b b67b  
RESULT 12  
AAT21884  
ID AAT21884 standard; cDNA to mRNA; 301 BP.  
XX AAT21884;  
XX  
AC AAT21884;  
XX  
DT 14-AUG-1996 (first entry)  
XX  
DE Human gene signature HUMGS03426.  
XX  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9514772-A1.  
XX  
PD 01-JUN-1995.  
XX  
PF 11-NOV-1994; 94WO-JP01916.  
XX  
PR 12-NOV-1993; 93JP-0355504.  
XX  
XX (MATSUBARA K.  
XX (OKUBO K.  
XX  
PI Matsubara K, Okubo K;  
XX  
DR WPI; 1995-206931/27.  
XX  
XX  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
XX  
PS Claim 1; Page 991-992; 2245PD; Japanese.  
XX  
XX A single-stranded DNA (or its complementary strand or the corresp.  
XX double-stranded DNA) which comprises one of the 7837 "GS" sequences  
XX given in AAT19001-T26837 and which is able to hybridize to part of  
XX human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
XX sequences were obtained from 3'-directed cDNA libraries prepared  
XX from various human tissues; synthesis of cDNA was initiated from the  
XX 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
XX untranslated sequence is unique to a particular mRNA species, almost  
XX all the 3'-oriented cDNAs hybridize with specific mRNAs. Each library  
XX is constructed so as to reflect accurately the relative abundance of  
XX different mRNAs in the particular tissue from which it was derived.  
XX The appearance frequency of a given GS in a cDNA library can be  
XX determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
XX  
SQ Sequence 301 BP; 96 A; 65 C; 63 G; 66 T; 11 other;  
Query Match 4.2%; Score 177.4; DB 16; Length 301;  
Best Local Similarity 79.3%; Pred. No. 3.9e-33;  
Matches 241; Conservative 0; Mismatches 57; Indels 6; Gaps 3;  
QY 3453 GATCAGATAGACACACAAAGAGAGAGATAGAGA--CCCTCGAGGAGAGTCCCT 3509  
DB 1 GATCAGATAGAGACACAG 60  
QY 3510 CTCGCCACCCCGCAGAGAGTCTCAACAGACACATTCCTGTATGCAAAAGAACCCAC 3569  
DB 61 CTCGCCACCCCGCAGAGAGTCTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
QY 3570 TAGCAGAGAGGCGCTGAGAGAGACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3629  
DB 121 TAGCAGAGAGGAG 179  
QY 3630 ACCACAG 3689  
DB 180 ACCACAG 239  
QY 3690 AGTTTAT 3749  
DB 240 AGTTTAT 297  
QY 3750 TACA 3753  
DB 298 TAAA 301  
RESULT 13  
ABA20255  
ID ABA20255 standard; DNA; 546 BP.  
XX  
XX ABA20255;  
XX  
AC ABA20255;  
XX  
DT 23-JUN-2002 (first entry)  
XX  
XX  
DE Human nervous system related polynucleotide SEQ ID NO 12586.  
XX  
XX  
KW Human; noctropic; neuroprotective; cytosolic; dermatological; virucide;  
KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnery;  
KW antiparkinsonian; antischistosomal; antianemic; antiarthritic; cancer;  
KW antileukemic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO200159063-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01334.  
XX  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.





CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_sequences.  
SQ Sequence 546 BP; 133 A; 137 C; 171 G; 105 T; 0 other;  
Query Match 4.1%; Score 171.4; DB 22; Length 546;  
Best Local Similarity 81.7%; Pred. No. 1.6e-31;  
Matches 223; Conservative 0; Mismatches 46; Indels 4; Gaps 2;  
QY 384 GGTGGAAGGAGTGAACACCCCAAGACCCAGAGTACCAACCAAGTCTCAGCCCATGGC 443  
D 2 GGTGGAAGGAGTGAACACCCCAAGACCCAGAGTACCAACCAAGTCTCAGCCCATGGC 61  
444 CCAAGACTTGGCCCTCAGATCTGTACCCCAATGAGACTTGGAGAAGCGAGTGAACC 503  
Db CCAAGACTTGGCCCTCAGATCTGTACCCCAATGAGACTTGGAGAAGCGAGTGAACC 121  
QY 504 CCAAGCTGAGAGAGGAGGCCAGCTGCAAGGCGAGAGGTTGGGCCCCAGCTGAAGAGA 563  
Db 122 CCAAGCTGAGAGAGGAGGCCAGCTGCAAGGCGAGAGGTTGGGCCCCAGCTGAAGAGA 180  
QY 564 GGG---AACTGAGAACCCCAAGAGCTTCAAGAGCTGTGAGAAATGGCTGTGTGAC 620  
Db 181 GGGTGCAGCTGAACCTCTGCTTGAAGCTTCAAGAGCTGTGAGAAATGGCTGTGTGAC 240  
QY 621 CAAGGAAGCGGTGAGAGCTTCTGCAAGAGAGG 653  
Db 241 CAAGGAAGCGGTGAGAGCTTCTGCAAGAGAGG 273  
RESULT 14  
AAS86166  
ID AAS86166 standard; cDNA; 2951 BP.  
XX  
AC AAS86166;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #21970.  
XX  
K human; chromosome mapping; gene mapping; gene therapy; forensic;  
K blood supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG21979.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnosis, forensic, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 1; SEQ ID No 21970; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantifying a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensic, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_sequences.  
XX  
SQ Sequence 2951 BP; 632 A; 850 C; 929 G; 540 T; 0 other;  
Query Match 4.1%; Score 169.8; DB 23; Length 2951;  
Best Local Similarity 45.9%; Pred. No. 9.3e-31;  
Matches 837; Conservative 0; Mismatches 952; Indels 36; Gaps 6;

QY 1133 TTGTGCTTGTGTGATGACAGCGCCGAGCCGAGCACTGAAGCACTGCTGGTCAATGT 1192  
Db 1082 TTGTGCTTGTGTGATGACAGCGCCGAGCCGAGCACTGAAGCACTGCTGGTCAATGT 1141  
QY 1193 GGTTCGAGATGGCAAGTTCTCAGTGTGTGTGTGTGAGAAAGCTCAGCCCTGAGTCT 1252  
Db 1142 GGTTCGAGATGGCAAGTTCTCAGTGTGTGTGTGTGAGAAAGCTCAGCCCTGAGTCT 1201  
QY 1253 TCTGCAATGATTCACAGGCGCCACTTACACAGAGCCCAATGATCCGGAAGCATCT 1312  
Db 1202 CTGAGGCTGTGCTGCTGCTGCTGCTTGTGATTAAGCTGCTGCTTGTGAAAGCATGT 1261  
QY 1313 ACGAAGTCTTCCAGGTGGCCAGAGCGCTGTGAGAGCCCAATGATCCGGAAGCATCT 1372  
Db 1262 ACCATGCTTGGAGGT-----AATGAGATGAGGGAATGAGGGCTTAAGCCTGAGAGC 1315  
QY 1373 GTGATGAAAGTGCAGTGGCAAGGCTGTGAAAGTGCAGAAACAGCATATGATGGG 1432  
Db 1316 AGGATGAAAGCAAGAACCCACAGAAATTGAGAGACCAAGCTGAAGCCCATGTTGAGTGGG 1375  
QY 1433 CCTCGGTGCTTCCAGCCCTCGGGTCTTAAGGGCTGTGAGCCAGCCAGAAAGAGAAGA 1492  
Db 1376 CCCACGGGGCTTCAAGCCCATCTGGATGAGAGGCTTCAAAACCAACACGACACAG 1435  
QY 1493 ATCTTACAAAGAAATTACACCGACATGTGGGTGAGAGCTGTAAGCAGCTTACGCCC 1552  
Db 1436 GTGGAAATGATGCCCATGACAGACCCGCTGCTGCTGTGGAGCACTTCTTGGGA 1495  
QY 1553 CACCCCAAGCCCAAGAAACCAAGAAAGCAACAGAGAAACTTAAGCTCAAGAGA 1612  
Db 1496 GAGTCC-----GCCTCAAGCAAAATTGCTATTAACAACGCGAAGACCGAGGGATGAGA 1550  
QY 1613 TCATTATGAGGCGACAGAGGAGCGGCTGTATGAGGGTGGCCAGAAAGTGCAGAAACA 1672  
Db 1551 TCAAGCCGAGGTGATTTGGGTACTTGGATTCATGGACAGATGGAGAGAGACGCTG 1610  
QY 1673 TCGAGACATTTGTATCTCATGTGGAGAGCTCAATGTCAACCTTGAGACACCACTTTCA 1732  
Db 1611 CAGATCAGGAATGATGATAT-----CCCTCAACCTCGGTGAGGAGCTTGAAGCC 1662  
QY 1733 TTGAGAGCATGTGCCGAAGTGTAAAGACTGTTCTTGGAGTGTGCTTACCAAGTATGACG 1792  
Db 1663 TGCCTTGTCTGACAGATGCGGGTAAAGTCTCTACTACTGCTTGAAGCTTCTCTCC- 1721

QY 1793 ACATGGGTACCACTCTATTGCAACATCTGCTGTGGGGGCGTGAAGTCTCATGTGTG 1852  
Db 1722 -CCTTGGCTCTCTAACTCCCTCTCTCTTCCCTGCGAGGCGCAGAGCTGCTGCTTTGCA 1780  
QY 1853 GGAACACAACTGCTGCGAGTGTCTTTTGTGTCAGTGTGGATCTCTTGTGTGGGGCCAG 1912  
Db 1781 GCAACACGAGCTGCTGCGGTGAGCACTGGGCGCTGTGGGTGGATGTGGGTGGGCCCC 1840  
QY 1913 GAGTGTCTAGGACGACATTAAGGAAGACCCCTGGAACCTGCTACATGTGGGGGATAAGG 1972  
Db 1841 AAGGCTCTTACGTTCTCGAGTCTAGGAGCCCTGGAGCTGTACATGTGTCTCCCGCAGC 1900  
QY 1973 GCACCTATGGGTGCTGCAAGAAGCGGAAGACTCGCCCTTCTGACCTCCAGATGTTCTTTG 2032  
Db 1901 GCTGTCTACGCTGCTGCGGCGCGAAGAGACTGGAACGTCGCCCTGCAAGGCTTTCTTCA 1960  
QY 2033 CCAATACCATGACACAGGAATTTGACCCCGCCCAAGGTTTATCCACCTGTGCCAGCTGAGA 2092  
Db 1961 CCACTGACACGGGCTTGAATATGTAAGCCACACAGGCTCCCGCTCTACCAACACAGATCC 2020  
QY 2093 AGAGGAAGCCATCCCGTGTCTCTCTTTGATGGATTGTCAGGGCTCTCTGGTGC 2152  
Db 2021 CAGGGGCACAGGCTGTTGAAAGCTCTGTTTGTATGTCATCGCGCAGGTGAGTTTCGGGG 2080  
QY 2153 TGAAGGACCTGGGATCCCAAGTGGACCGCTACATTGCTCCGAGGTGTGAGGACTCCA 2212  
Db 2081 AACACCTGGAGACATGCTATCGTGTCAACACAGGCTAGCCAGGAGTCAAGAGTCCA 2140  
QY 2213 TCAGGTGGGCTGTGGCGGACAGGGAAGATCATGTACGTGCGGAGCTCCGCGAGG 2272  
Db 2141 TTGCTGTGGAACCGTGAAGCAGGAGGGAATATCAATACGTGAACGACGTGAGGAACA 2200  
QY 2273 TCACACAGAAGCATATCCAGGAGTGGGGCCCATTCGACCTGCTGTGATGGAGCAGTCCCT 2332  
Db 2201 TCACAAAGAAAATGTGAGGGCAGTCTGTACTTTGCGGGGCTCATCTCTCTGCTGCTTT 2260  
QY 2333 GCAATGACCTCTCCATTGTCAACCTGCGCCCAAGGACATTTATG-----AGGGAATG 2386  
Db 2261 TCCCGAGTCTCCACACCCGAATCCAGCAGGAAGGCGCTGTATGTTGAGCATCTCTTCTC 2320  
QY 2387 GCCGCTCTTCTTTGAGTGTCTACCCGCTCTCTG-----CATGATGCGCGGCCAAGG 2437  
Db 2321 TCTGGCAGTCTCTGGAGAGCTTATGTACCTGACCTGACCTGCCCGAGGCGAGGAATAAG 2380  
QY 2438 AGGAGATGATGCCCTTCTTCTGCTCTTTTGAGAATGTGTGGCCATGGCGCTTAGTG 2497  
Db 2381 TGGGTTGGGCTGGAGGTTTCAATGAACCTCGCTGCTCATCTTTTCTGAGCACAGAGG 2440  
QY 2498 ACAGAGGAGCATCTCGGATTTCTTGAGTCTNAACCCGCTGATGATTGAGCCCAAGAG 2557  
Db 2441 TACAGAGGACATCTCACGGTTCTTGGAGGTGAGGGAATCTGGGACCTGATTGTCAAG 2500  
QY 2558 TGTCTGTGCACACAGGCGCGTTACTTCTGGGGTAACTTCTGGGATGAACAGGCGCTT 2617  
Db 2501 ACAGCCAGGCGAGGAAAGCGCTGTGTCAGTGAACCTTACCCGGGATGAACAGGTAAC 2560  
QY 2618 TGGCATCTGTAATGATAAGCTGGAGCTGCAAGAGTGTCTGGAGCAGCGCAAGATAG 2677  
Db 2561 AAAGGCTCTTAGTGGGTGAGTAAACAGCCAAAGTTAAATATGTGATAACAAAGCTTGACA 2620  
QY 2678 CCAAGTTCAAGAAAGTGAAGACCATTAACACAGGTCACAACTCTATAAAGCAGGCAAG 2737  
Db 2621 CCAAGGTAAGACAGCTGTGGCGCTCTGGAAAAATGCACTTGGTACCTCCCAAGTGGGA 2680  
QY 2738 ACCAGCATTTCCCGCTTTCATGAACGAGAGGAGGACCTCTGTTGTCATGAAATGG 2797  
Db 2681 CTTGGGAGATCCCTGTTGTATGATGAATGCAAGAAAGATGTTTGTGTGCACTGAGCTCG 2740  
QY 2798 AAAGGCTGTTGCTCCCGTCCACTACACAGAGCTCTCAACATGAGCGCTTGGGGA 2857  
Db 2741 AAAGGTAGCAAGGCTGCACTTGGAGAGGGAACCTGTGTAGATCAAAACACAAATGGCA 2800  
QY 2858 GGCAGACATGCTGGGCCCATCTGTGGAGCGTGGCGGTATCCCGCACCTCTTTCGCTCCG 2917

Db 2801 GACATGGGCTGCGGAAGGTCCTGGAGGCTGCTGTATCCGACACCTCTTCGCCCTC 2860  
QY 2918 TGAAGGAATATTTGCTTGTGTGTA 2942  
Db 2861 TGAAGGACTACTTTTGCATGTGAATA 2885  
RESULT 15  
ABAI3731  
ID ABAI3731 standard; cDNA; 283 BP.  
XX ABAI3731;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Human nervous system related polynucleotide SEQ ID NO 2738.  
XX  
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antispasmodic; antianemic; antiarthritic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;  
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.  
XX  
OS Homo sapiens.  
XX  
WO200159063-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01334.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220863.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224318.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.



Db 141 AATGGGACTTGGAGAAGCGAGTGAGCCCCAGCCAGAGGAGGGAGCCCTGCTGGGGG 200  
Qy 535 CAGAGGGTGGGGCCCCAGCTGAAGGAGGG---AACTGAGACCCCAAGAGCCTCC 591  
Db 201 CAGAGGGCGGGCCCCAGCAGAGGAGGGGTGACCTGACCTGCTGAGCCCTCA 260  
Qy 592 AGAGCTGTGAGAAATGGCTGCTG 614  
Db 261 AGAGCAGTGGAAAATGGCTGCTG 283

Search completed: November 22, 2002, 09:09:32  
Job time : 605.945 secs

[illegible]

THIS PAGE BLANK (USPTO)

QY 321 GILYVTHrPheSerSerProGlyGluSerLeuGluAspGluLeuLysProMetLeu 340  
DB 1229 GGCAAGACCTTCTCCAGAGCTCTGGAGAGTCACTGAGGAGCACTGAAGCCATGCTG 1288  
QY 341 GILYTPAlaHISGLYGLYPhLeuLysProThrGlyIleGluGlyLeuLysProAsnLys 360  
DB 1289 GAGTGGGCCCCAGGTGGCTTCAAGCTTACCTGAGATCGAGGCGCTCAACCAAGAG 1348  
QY 361 GILYProValValAsnLysSerValArgArgSerSerSerArgAsnLeuGluProArg 380  
DB 1349 CAACCACTGGTTAATAGTGAAGTCTGCTGCTTCAAGAGTGAAGTGAAGTGAAG 1408  
QY 381 ArgArgGluAsnLysSerArgArgThrThrAsnAspSerAlaIleSerGluSerPro 400  
DB 1409 AGACGGAGAACAAAGTGGAGACCAACCAATGACTCTCTCTGAGTGGTCCCC 1468  
QY PropGlyAsArgLeuLysThrAsnSerValGlyGlyLysAspArgGlyLysAspGluGlu 420  
DB CCACCCAGCCCTCAAGACCAATAGCTATGCGGAGAGACCGAGGGAGATGAGAG 1528  
QY 421 SerArgGluArgMetAlaSerGlyValThrAsnLysGlyAsnLeuGluAspArgCys 440  
DB 1529 AGCCGAGAACCGATGGCTTCTGAGTCAACCAAGAGGCAATCTGAGAGACCGCTGT 1588  
QY 441 LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyLysCys 460  
DB 1589 TTGCTCTGTGAGAAAGAAACCTGTGCTTCCACCCCTCTTGGAGGGGCTGCTGT 1648  
QY 461 GluSerCysArgAspArgPheLeuGluLeuPheThrMetTyrAspGluAspGlyTyrGln 480  
DB 1649 CAGAGTGGCGGAGTCTCTCTTCAAGCTCTTCAATGATGAGAGCGCTATCAG 1708  
QY 481 SerTyrCysThrValCysCysGluGlyArgGluLeuLeuLysSerAsnThrSerCys 500  
DB 1709 TCTTACGACCTGTGCTGTGAGGCGCTGAACCTGCTGTGAGTGAACACAGCTGC 1768  
QY 501 CysArgCysPheCysValGluCysLeuGluValLeuValGlyValGlyThrAlaGluAsp 520  
DB 1769 TCCAGATGCTCTGTGTGAGTGTGAGGCTGTGAGGCTGTGAGGCGAGAGTGAAGT 1828  
QY 521 AlaLysLeuGluGluProThrSerCysTyrMetCysLeuProGluArgCysHisGlyVal 540  
DB 1829 GCCAAGCTGCAAGAACCTGTGAGTGTATGTCCTCTCAAGCTGTGAGGCTGTGAG 1888  
QY LeuAlaGlyGlyValAspTyrAsnMetArgLeuGluAsnPheThrThrAspProAsp 560  
DB CTCCAGCCAGAGAAAGATTGAACTGCGCTGCAAGACTTCTCACTACGATCTCTGAC 1948  
QY 561 LeuGluGluPheGluProProLysLeuTyrProAlaIleProAlaIleLysArgArgPro 580  
DB 1949 CTGAGAGATTTGAGCCACCAAGTTGACCCAGCAATTCCTGAGCCAAAGAGAGCCC 2008  
QY 581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeu 600  
DB 2009 ATTAGAGTCTCTGCTCTGTTGATGGAATTGCAACCGGAGTACTTGGTCTCAAGAGTTG 2068  
QY 601 GlyIleLysValGluLysTyrIleAlaSerGluValCysAlaGluSerIleAlaGly 620  
DB 2069 GGATATTAAAGTGAAGAAAGTATGCTGCTCCGAACTGTGAGAGTCAATCGCTGGGA 2128  
QY 621 ThrValLysHisGluGlyGlnIleLysTyrValAsnAspValArgLysIleThrLysLys 640  
DB 2129 ACTGTTTAAAGCATGAAGGCGCAGATCAAAATATGTCATGACGTCGGAAAAATCACCAAGAA 2188  
QY 641 AsnIleGluGluTyrGlyProPheAspLeuValIleGlyGlySerProCysAsnAspLeu 660  
DB 2189 AATATGAAGAGTGGGCGCTGCTGACTTGTGATGTTGGTGAAGCCCATGCAATGATCTC 2248  
QY 661 SerAsnValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhePheGlu 680  
DB 2249 TCTAAAGTCATCTGCGCGCAAGGTTTATATGAGGACAGAGAGGCTCTTCTGAG 2308  
QY 681 PheTyrHisLeuLeuAsnTyrThrArgProLysGluLysAspAsnArgProPhePheTyr 700

DB 2309 TTTTACCACTGCTGANTTATACCCGCCCCCAAGAGGCGCACAACCGCTTCTTCTG 2368  
QY 701 MetPheGluAsnValValAlaMetLysValAsnAspLysLysAspIleSerArgPheLeu 720  
DB 2369 ATGTTGAGATGTTGTGGCCATGAAGTGAATGACAAAGAAAGCATCTCAAGATTCTG 2428  
QY 721 AlaCysAsnProValMetIleAspAlaIleLysValSerAlaIleArgAlaArgTyr 740  
DB 2429 GCATGTATCCCAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 2488  
QY 741 PheTyrGlyAsnLeuProGlyMetLysAsnArgProValMetAlaSerLysAsnAspLys 2548  
DB 2489 TTCTGGGTTAACTTACCCGGAATGAACAGGCGCGTATGCTTCAAAAGATATAGCTC 2608  
QY 761 GILYLeuGluAspCysLeuGluPheSerArgThrAlaLysLeuLysValGlnThrIle 780  
DB 2549 GAGCTGACAGACTGCTGCTGAGTTCACTAGACAGACGAAAGTTTAAAGAAAGTCA 2608  
QY 781 ThrThrLysSerAsnSerIleArgGlnLysAsnGluLeuPheProValValMetAsn 800  
DB 2609 ACCACCAAGTCAAACTCCATCAGACAGGCGCAAAACCACTTTTCTGTTAGTCAAT 2668  
QY 801 GlyLysAspAspValLeuTyrCysThrGluLeuGluArgIlePheGlyPheProAlaHis 820  
DB 2669 GGCAAGAGCAGCTTTTGTGTGTGATGATGATGATGATGATGATGATGATGATGAT 2728  
QY 821 TyrThrAspValSerAsnMetGlyArgGlyAlaArgGlnLysLeuGluLysArgSerTyr 840  
DB 2729 TACACGACCGTGTCCAAACATGAGCGCGGCGCGCGCTGTAGAACTGTGAGCAGCTCG 2788  
QY 841 SerValProValIleArgHisLeuPheAlaProLeuLysAspTyrPheAlaCysGly 859  
DB 2789 AGGTACCGCTATCAGACACTGTTGGCCCCCTTGAAGACTACTTCTGCTGAA 2845

RESULT 2  
ARIS1973  
LOCUS

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 22, 2002, 03:58:32 ; Search time 4894.5 Seconds

(without alignments)  
5107.633 Million cell updates/sec

Title: US-09-720-086-6

Perfect score: 4590  
Sequence: 1 MKGDSRHINEEGASGYEEC.....MSVPVIRHLPALPDYFACE 859

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool/US09720086/runat.18112002.092832.22195/app.query.fasta\_1.4252  
-DB=genembl -QPM=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09720086.CGN.1.11308 @runat.18112002.092832.22195 -NCPU=6 -ICPU=3  
-NO.XLPHY -NO.MAP -LARGEDURY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-MARY.TIMEOUT=30 -THRAADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Data base :  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pt:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vt:\*  
15: gb\_vt:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_mu:\*  
21: em\_mu:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*

29: em\_vt:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vit:\*  
38: em\_sy:\*  
39: em\_higo\_hum:\*  
40: em\_higo\_mus:\*  
41: em\_higo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4590	100.0	4195	10 AF068626	AF068626 Mus muscu
2	4573	99.6	4223	10 AF151973	AF151973 Mus muscu
3	4573	99.6	4338	10 AF151969	AF151969 Mus muscu
4	4472	97.4	4135	10 AF068627	AF068627 Mus muscu
5	4455	97.1	4163	10 AF151974	AF151974 Mus muscu
6	4455	97.1	4278	10 AF151970	AF151970 Mus muscu
7	4223.5	92.0	4006	10 AY078427	AY078427 Mus muscu
8	4206.5	91.6	4034	10 AF151975	AF151975 Mus muscu
9	4206.5	91.6	4149	10 AF151972	AF151972 Mus muscu
10	4105.5	89.4	3946	10 AF068628	AF068628 Mus muscu
11	4088.5	89.1	3974	10 AF151976	AF151976 Mus muscu
12	4088.5	89.1	4089	10 AF151971	AF151971 Mus muscu
13	3759	81.9	4145	9 AF164488	AF164488 Homo sapi
14	3759	81.9	4335	9 AF151857	AF151857 Homo sapi
15	3675	80.1	4267	9 AF176228	AF176228 Homo sapi
16	3311.5	72.6	3897	9 AF164487	AF164487 Homo sapi
17	2453	53.4	3017	9 AK001191	AK001191 Homo sapi
18	2206.5	48.1	3005	9 AF067972	AF067972 Homo sapi
19	2206.5	48.1	4258	9 AF331856	AF331856 Homo sapi
20	2185	47.6	4192	10 AF068625	AF068625 Mus muscu
21	2184	47.6	4094	10 BC007466	BC007466 Mus muscu
22	1832.5	39.7	2848	5 AF135438	AF135438 Danio rer
23	1734.5	37.8	2191	9 AK025230	AK025230 Homo sapi
24	1428.5	31.1	17697	2 AC112586	AC112586 Rattus no
25	1352	29.5	123936	2 AC1120824	AC1120824 Rattus no
26	1271	27.7	167568	2 AC111734	AC111734 Rattus no
27	1216	26.5	244329	2 AC107644	AC107644 Mus muscu
28	1209	26.3	1758	9 BC018214	BC018214 Homo sapi
29	1183	25.8	16851	9 AC009474	AC009474 Homo sapi
30	1128	24.6	2127	9 AF129267	AF129267 Homo sapi
31	987.5	21.5	2057	9 AF129268	AF129268 Homo sapi
32	882	19.2	2008	9 AF129269	AF129269 Homo sapi
33	862	18.8	2077	6 AR129189	AR129189 Sequence
34	813	17.7	225045	2 AL833803	AL833803 Mus muscu
35	811.5	17.7	249245	2 AC123256	AC123256 Mus muscu
36	695.5	15.2	18936	2 AC128509	AC128509 Rattus no
37	675.5	14.7	1705	9 AF194032	AF194032 Homo sapi
38	673	14.7	1397	9 BC002560	BC002560 Homo sapi
39	660	14.4	1589	10 AF220524	AF220524 Mus muscu
40	660	14.4	1645	10 MM040467	MM040467 Mus muscu
41	632.5	13.8	110000	9 AL354832	AL354832 Homo sapi
42	632.5	13.8	118899	9 HS1085F17	HS1085F17 Human DNA
43	618	13.5	210269	2 AC116459	AC116459 Mus muscu
44	597.5	13.0	119630	2 AC112040	AC112040 Rattus no
45	571	12.4	153533	2 AC095199	AC095199 Rattus no

RESULT 1

ALIGNMENTS



QY 321 GlyThrPheSerSerProGlyLeuGlnLeuPheProMetLeu 340  
 Db 1229 GGCAAGACCTTCTCCAGAGCTCTGAGAGTCACGAGGACCACTGAAGCCCTGCTG 1288  
 QY 341 GluTPAlaHisSGLYGLYPheLeuProThrglyLeuGlyLeuLeuProAsnLeuSls 360  
 Db 1289 GAGTGGGCCACGGTGGCTTCAAGCTTACTGGGATCGAGGGCTTAAACCCCAAGAG 1348  
 QY 361 GlnProValAlaAsnLeuSerLeuValArgArgSerAspSerArgAsnLeuGluProArg 380  
 Db 1349 CAACAGATGTTAATATAGTCGAGAGTGCGTCTTCAAGACAGTGAACCTTGAACCCAGG 1408  
 QY 381 ArgArgGluAsnLeuSerArgArgThrThrAsnAspSerAlaAlaSerGluSerPro 400  
 Db 1409 AGACGGCGAACAAGTCGAGAGCGCAACCAATGACTCTGCTTCTGAGAGTCCCC 1468  
 QY 401 ProProValSGLYLeuValThrAsnSerThrglyLeuAspArgGlyGluAspGlyLeu 420  
 Db 1469 CACCCAGCGCTTCAAGACAAATAGCTATGCGGGAGAGACCGAGGGAGAGTAGAGAG 1528  
 QY 421 SerArgGluArgMetAlaSerGlyValThrAsnAsnLeuSGLYAsnLeuGluAspArgCys 440  
 Db 1529 AGCGAGAACCGATGGCTTCTGAATCAACCAACAGGCAATCTGGAAGACCGCTGT 1588  
 QY 441 LeuSerCysGlyLeuValAsnProValSerPheHisProLeuPheGlyGlyLeuCys 460  
 Db 1589 TTGTCCTGTGAGAAAGAAACCTGTGTCTTCCACCCCTCTTGAAGGTTGGCTGT 1648  
 QY 461 GlnSerCysArgAspArgPheLeuGluLeuPheThrmetyrAspGluAspGlyLeuGln 480  
 Db 1649 CAGAGTTCGGGATGCTCTTCTGAGCTCTTCAATGATGATGAGAGACCGCTATAG 1708  
 QY 481 SerThrgThrValCysCysGlyGlyValArgGluLeuLeuLeuCysSerAsnThrSerCys 500  
 Db 1709 TCCATCTGACCGTGTGCTGTGAGGGCCGTGAACCTGCTGTGAGTAAACAAGCTGC 1768  
 QY 501 CysArgCysPheCysValGluCysLeuGluValLeuValGlyAlaGlyThrAlaGluAsp 520  
 Db 1769 TGCAATCTCTCTGTGTGAGAGTGTGTGAGAGTGTGTGTGGCGCAGCAGCTAGAGAT 1828  
 QY 521 AlaLeuLeuGlnGluProTrpSerCysThrmetyrCysLeuProGlnAspCysHisSGLYAla 540  
 Db 1829 GCCAAGCTGACGAAACCTGTGAGCTGTATATGCTCTCCACGCGTCCATGGGCTC 1888  
 QY 541 LeuArgGAGLYAspTrpAsnMetArgLeuGlnAspPhePheThrThrAspProAsp 560  
 Db 1889 CTCGACCCAGGAAGATTTGGAACATGGCCCTGCAGAGCTTCTTACTATGATCTTAC 1948  
 QY 561 LeuGlnGluPheGluProProLeuLeuThyrProAlaIleProAlaAlaLeuArgArgPro 580  
 Db 1949 CTGGAAGATTTGAGCCACCCAAAGTTGACCCAGCAATTCCTGCGCAAAAGAGGCC 2008  
 QY 581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyThyrLeuValLeuLeuSGLYLeu 600  
 Db 2009 ATTAAAGTCCGTCTCTGTGTGATGAAATGCAACGGGGTACTTGTGTCTCAAGAGTTG 2068  
 QY 601 GlyIleLeuValGluLeuThyrIleAlaSerGluValCysAlaGluSerIleAlaValGly 620  
 Db 2069 GGTATTAAGTGAAGAAAGTACTTGCCTCCGAGTCTGTGCAAGTCCATCTCTGGGA 2128  
 QY 621 ThrValLeuHisSGLYGLYLeuLeuThyrValAsnAspValArgLeuIleThrLeuSls 640  
 Db 2129 ACTGTTAAGCATGAAGGCCAGATCAATATGTCAATGACGTCCGGAATAATCCCAAGAA 2188  
 QY 641 AsnIleGluGluThrglyProPheAspLeuValIleGlyGlySerProCysAsnAspLeu 660  
 Db 2189 AATATTTGAAGAGTGGGGCCGCTTCACTGTGATTTGGGAGGCCATGCAATGATCTC 2248  
 QY 661 SerAsnValAsnProAlaArgLeuSGLYLeuThrglyValArgLeuPheGlu 680  
 Db 2249 TCTAACGTCATCTCTGCGCCGCAAGGTTTATATGAGGGCAGCAAGAGGCTTCTTCCAG 2308  
 QY 681 PheThrHisLeuLeuAsnThyrThrArgProGlySGLYAspAsnArgProPhePheTrp 700

Db 2209 TTTTACACCTTGTGAAATATATCCGGCCCAAGGAGCGCAACCGTCTTCTTCTGG 2368  
 QY 701 MetPheGluAsnValAlaAlaMetLeuValAsnAspLeuSlsAspIleSerArgPheLeu 720  
 Db 2369 ATGTTCAAGAAATGTTGGCCATGAAGATGATGACAAAGAAACATCTCAAGATTCTG 2428  
 QY 721 AlaCysAsnProValMetIleAspAlaIleLeuValSerAlaAlaHisArgAlaArgTyr 740  
 Db 2429 GCATGTAACCAAGATGATGATGATGCCATCAAGGTGTCTGCTGTCAAGGGCCGGTAC 2488  
 QY 741 PheTrpGlyAsnLeuProGlyMetAsnArgProValMetAlaSerLeuSlsAsnAspLeu 760  
 Db 2489 TTTGGGGTAACTTACCCGGAATGAACAGCGCGGTGATGCTTCAAGAAATGATTAACCTC 2548  
 QY 761 GlnLeuGlnAspCysLeuGluPheSerArgThrAlaLeuLeuSlsValGlnThrIle 780  
 Db 2549 GAGCTGCAGACTGCTGTGAGTCAAGTACGACCAAGTTAAAGAAAGTCAAGACATA 2608  
 QY 781 ThrThrLeuSerAsnSerIleArgGlnGlyLeuAsnGlnLeuPheProValAlaMetAsn 800  
 Db 2609 ACCACCAAGTGAATCTCCATCAGACAGGGCAAAACAGCTTTCCCTGTATGATGAAT 2668  
 QY 801 GlyLeuAspAspValLeuTrpCysThrglyLeuGluArgIlePheGlyPheProAlaHis 820  
 Db 2669 GGCAAGCGACGTTTGTGTGTGCTGACCTGAGCTCGAAGATCTTCGCTTCCCTGCTCAC 2728  
 QY 821 ThyrThrAspValSerAsnMetGlyArgGlyValAlaArgGlnLeuLeuGluArgSerTrp 840  
 Db 2729 TACACGAGCTGTCCAAATAGGCGCGGCGCCGTCAGAAAGTGTGTGGAGATCTTGG 2788  
 QY 841 SerValProValIleArgHisLeuPheAlaProLeuLeuAspThyrPheAlaCysGlu 859  
 Db 2789 AGTGAACGGTCATCAGACACCTGTTGGCCCTTGAAGAGCATCTTGTCTGTGAA 2845  
 RESULT 2  
 AF151973 4223 bp mRNA linear ROD 08-JUN-2000  
 LOCUS AF151973  
 DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 5  
 (Dnmt3b) mRNA, complete cds.  
 ACCESSION AF151973  
 VERSION AF151973.1 GI:8347127  
 KEYWORDS  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 4223)  
 Yln,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and  
 Shen,Y.  
 Cloning of full-length Dnmt3b cDNA and its alternative splicing  
 isoforms in mouse embryonic tissue  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 4223)  
 AUTHORS Yln,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and  
 Shen,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular  
 Biology, Institution of Basic Medical Sciences, Chinese Academy of  
 Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,  
 P.R.China  
 FEATURES  
 source Location/Qualifiers  
 1..4223  
 /organism="Mus musculus"  
 /strain="KM"  
 /db\_xref="taxon:10090"  
 /dev\_stage="8-9 day old embryo"  
 1..4223  
 /gene="Dnmt3b"  
 /gene="Dnmt3b"  
 297..2876  
 /function="methylates cytosine in DNA"  
 /note="alternatively spliced"

```

/codon_start=1
/product="DNA cytosine-specific methyltransferase isoform
5"
/protein_id="AAP74519.1"
/db_xref="GI:8347128"

```

```

/translations="MGDSRLHNEEGASGYEBGIIIVNGNFSQSDSDTKDAPSPVLE
ALCTPEVCTETGRSSRLSRVSSLLAYTQMDTGDGRDDEVDGSDILMPK
LTRKETRTRESAPVTRTHNSGTSLEKQASPRITRGRQGHVQBPVPEFPAIR
SRRRASSASTPWSPASVDMEEVTPKSVSTPSVLSQDQDQDMDTQVDAESIY
GDEYQDDKDFGIGDLVWGKIKFSGWMPAMVSVKATSKRQAMPGRWQMGDFK
SEISADLVALGLFSGHFNLATFNKLVSRKAMYHTLEKARVRAGTKFFSSPGSELE
QKAPLMAHAGGPKPTGIEGLKPNKKQPVVANKSVKRSRDLPRRRKSRRTTN
DAASESPPPKRLKNTSYGCKORDEESRERMASEVTNNKGNLDCSLSCCKQNPVS
FPHLEGLQSCDRDRLFLFYMYDEYQSYCTVCCGELLCSNTSCRCFCVEP
LEVLVGAPAAKRAPIRVLISLFDGIATGLYLKELGIVKERYIASEVCAESIAVGVKH
EGQIKYVNDVRKITTNNIEBWPFDLVIGGSPCNLDSNVNPAKGLYEGTGRLEPEFY
HLNATPXPBGDNRPFFMNFENVAMKVDKDI SRFLACNPVMI DAIKVSAAHRAHY
FMGLPKMNPVMAKNDKLEQDCLBEFSRTAKLKVKQITITTKNSNIROGKNLPVV
MNGKDDVLWCTELERIFGPPAHYTDVSNMGRQAKLGRSVSPVIRHLFAPLKDYE
ACE"

```

BASE COUNT 1051 a 1084 c 1147 g 941 t

ORIGIN

#### Alignment Scores:

```

Pred. No.: 1,92e-248 Length: 4223
Score: 4573.00 Matches: 857
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 2
Query Match: 99.63% Indels: 0
DB: 10 Gaps: 0

```

US-09-720-086-6 (1-859) x AF151973 (1-4223)

```

Qy 1 MetLysGlyAspSerArgHisLeuAsnGluGluGluGlyValaserGlyTYrGluGluCys 20
Db 297 ATGAAGGGAGACAGACAGACATCTCAATGAAGAAGAGGGTGCAGCGGGTATGAGAGTGC 356
Qy 21 IleIleValAsnGlyAsnPheserAspGlnSerSerAspThrLysAspAlaProSerPro 40
Db 357 ATTATCGTTAATGGGAACCTTCAGTGACCACTCTCAGACACAGGAAGTGTCTCCCTCACCC 416
Qy 41 ProValLeuGluAlaIleCysThrGluProValCysThrProGluThrArgGlyValArg 60
Db 417 CCAAGTCTTGGAGGCAATCTGCACAGAGCCAGTCTGCACACACAGACACAGAGCGCGAGG 476
Qy 61 SerSerArgLeuSerLysArgGluValSerSerLeuAsnTyThrGlnAspMet 80
Db 477 TCAAGCTCCCGGCTGTCTAGAGGGAGGTCTCCAGCTTCTGAAATTACACAGGACATG 536
Qy 81 ThrGlyAspGlyAspArgAspGluValAspAspGlyAsnGlySerAspIleLeuMet 100
Db 537 ACAGGAGATGGAGACAGAGATGATGAAGTAGATGATGGGAATGGGCTCTGATATTCTAATG 596
Qy 101 ProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaValArg 120
Db 597 CCAAAGCTCACCGGTGAGACCAAGACACACAGGACGCGCTCTGAAAGCCCGGCTGTCCGA 656
Qy 121 ThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArgAlaserProArgIleThr 140
Db 657 ACCCGACATAGCAATGGGACCTCCAGCTTGGAGAGCAAGAGCTCCCTCCCCAGAAATCAC 716
Qy 141 ArgGlyArgGlnGlyArgHisGluValGlnGluTyProValGluPheProAlaThrArg 160
Db 717 CGAGGTTCGACAGGGCGGCCACCATGTGCAGAGTACCTCTGTGGAGTTTCCGGCTACCAAG 776
Qy 161 SerArgArgArgArgGlnSerSerAlaserThrProTrpSerSerProAlaSerVal 180
Db 777 TCTCGAGACGTTCAGACATCGTCTTCAGCAAGCACCCCATGTCATCCCTTCCCGCGCTC 836
Qy 181 AspPheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGln 200
Db 837 GACTTCATGGAAGAAGTAGACACCTTAAGAGCGTCAGTACCTACCTACTACTACTACTGAC 896

```

```

Qy 201 AspGlyAspGlnGluGlyMetAspThrThrGlnValAspAlaGluSerArgAspGlyAsp 220
Db 897 GATGAGATCAGGAGGTATGATATCCACACAGGTGATGATGAGAGCATATATGAGAC 956
Qy 221 SerThrGluTyThrGlnAspAspLysGluPheGlyIleGlyAspLeuValTrpGlyLysIle 240
Db 957 ACCACAGATATCAGATATGAAGATTGGAATAGTAGTACCTCGGTGTGGGGAAGATC 1016
Qy 241 LysGlyPheSerTrpTrpProAlaMetValSerTrpLysAlaThrSerLysArgGln 260
Db 1017 AAGGGCTTCTCGGTGGCTCCCATGGTGGTGTCTCTGGAAAGCCACCTCCAAGCGACAG 1076
Qy 261 AlaMetProGlyMetArgTrpValGlnTrpPheGlyAspGlyLysPheSerGluIleSer 280
Db 1077 GCCATGCCCGGAATGCGCTGGGTACAGTGGTGTGGTATGCGAAGTTTCTCAGATCTCT 1136
Qy 281 AlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsn 300
Db 1137 GCTGACAAACTGGTGGCTCTGGGGCTGTTCAGCCAGCACCTTTAATCTGCTACCTTCAAT 1196
Qy 301 LysLeuValSerTyArgLysAlaMetTyRHisThrLeuGluLysAlaArgValArgAla 320
Db 1197 AAGCTGGTTCCTTATAGGAAGCCATGTACACACTCTGGAGAAAGCCAGGTTCAGCT 1256
Qy 321 GlyLysThrPheSerSerProGlyLysLeuGluAspGlnLeuLysProMetLeu 340
Db 1257 GGCAGAGACCTTCTCCAGCAGTCTCTGGAGATCCTGGAGGACGAGCTGAAGCCCATGCTG 1316
Qy 341 GluTrpAlaHisGlyGlyPheLysProThrGlyIleGluGlyLeuLysProAsnLysLys 360
Db 1317 GAGTGGGCCACAGTGGTCTCAAGCTCTACGGATCGAGGGCTCAAAACCCCAACAGAG 1376
Qy 361 GlnProValValAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProArg 380
Db 1377 CAACAGTGGTTAATAGTCAAGGTGCGTCTGTTCAGACAGTAGTAGAACTTAGAACCCAGG 1436
Qy 381 ArgArgGluAsnLysSerArgArgThrThrAsnAspSerAlaAlaSerGluSerPro 400
Db 1437 AGACGCGAGAAACAAAGTCCGAGACGCAACCAATGACTCTGCTCTCTGAGTCCCCC 1496
Qy 401 ProProLysArgLeuLysThrAsnSerTyGlyLysAspArgGlyGluAspGluGlu 420
Db 1497 CCACCCAGCGCTCAAGCAAAATAGCTATGCGGGAGAGCCGAGGGGAGGATGAGGAG 1556
Qy 421 SerArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCys 440
Db 1557 AGCCGAGAACCGATGCGTCTCTGAAGTCCAAACCAAGGGCAATCTGGAAGACCGCTGT 1616
Qy 441 LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyLysLeuCys 460
Db 1617 TTGTCCTGTGGAAGAAAGAACCTGTGTCTTCCACCCCTCTTTGAGGGTGGGCTCTGT 1676
Qy 461 GlnSerCysArgAspArgPheLeuGluLeuPheTyRMetTyRAspGluAspGlyTyRGl 480
Db 1677 CAGAGTTGCCGGGATCGCTTCTTAGAGCTCTTCTACATGTATGATGAGGACGGCTATCAG 1736
Qy 481 SerTyRThrValCysCysGluGlyArgGluLeuLeuLeuCysSerAsnThrSerCys 500
Db 1737 TCCTACTGCACCGTGTGCTGTGAGGCGCGTGAACCTGCTGTGTGCAAGTAACAAGTGC 1796
Qy 501 CysArgCysPheCysValGluCysLeuGluValLeuValGlyAlaGlyThrAlaGluAsp 520
Db 1797 TCGAGATGCTTCTGTGTGGAGTGTCTGGAGGTGTGTGGGCGCAGGACACAGCTGAGAT 1856
Qy 521 AlaLysLeuGlnGluProTrpSerCysTyRMetCysLeuProGlnArgCysHisGlyVal 540
Db 1857 GCCAAGTTCAGCAAGAACCTCTGGAGCTGTCTATGTGCTCTCCCTCAGCGCTGCGCATGGGTC 1916
Qy 541 LeuArgArgArgLysAspTrpAsnMetArgLeuAsnAspPhePheThrThrAspProAsp 560
Db 1917 CTCGACCGAGGAAGATTGGAAACATGCGCTGCAAGACTTCTTCACTACTACTACTGATCCTGAC 1976

```

Qy 561 leuGlugluPhegluProProlybLeuTyrProAlaIleProAlaAlaIleAlaArgArgPro 580  
 Db 1977 CTGAAGAAGATTGAGCCACCCAGATTGACCCAGCAATTCCTGAGCCAAAGAGAGGCC 2036  
 Qy 581 lleArgValleuSerleuPheaspGlylleAlaThrGlyTyrleuValleuLeuGlyleu 600  
 Db 2037 ATTGAAGTCTCTCTCTCTCTGATGAGATTGCAACGGGGTACTTGCTGCTCAAGAGATTG 2096  
 Qy 601 GlylleuValGlyuTyrrlleAlaSerGluValCysAlaGluSerlleAlaValGly 620  
 Db 2097 GGTATTAAAGTGAAGAAAGTACATTGCTCCGAGCTGTGTGACAGTCCATCGCTGGGA 2156  
 Qy 621 ThrVallyshisgluGlyGlylleGlyTyrValIleAspValArgGlylleThrLysLys 640  
 Db 2157 ACTCTTAAGCATGAAGGCCAGATCAAAATATGTCAATGACGTCCCGAAATACCAAGAAA 2216  
 Qy 641 AsnIleGlugluTyrPglProPheAspIleValIleGlyGlySerProCysAspAspIleu 660  
 Db 2217 AATATTGAAGAGTGGGGCCCTTGACCTTGATGTTGGTGGAGGCCATGCAATGATCTC 2276  
 Qy 661 SerAsnValaenProAlaArgLysGlyLeuTyrGlyGlyThrGlyArgleuPhePheglu 680  
 Db 2277 TCTAAGCGCATCTCGCCCGGAAAGTTATATAGAGGCAAGAGGCTCTTCTTGAG 2336  
 Qy 681 PheTyrHisleuLeuAsnTyrThrArgProlysglyAspAsnArgProPhePheTyr 700  
 Db 2337 TTTTACCACTTGCTGAATTATATACCCGCCCAAGAGGGCGCAACCCGTCATCTTCTCG 2396  
 Qy 701 MetPhegluAsnValaIleAlaMetLysValaAspLysAspIleSerArgPheleu 720  
 Db 2397 ATGTTGGAAGATGTGGCCATGAGAAATGATGCAAGAAAGCATCTCAAGATCTCTG 2456  
 Qy 721 AlaCysAsnProValMetIleAspAlaIleLysValSerAlaAlaIleAlaArgAlaArgTyr 740  
 Db 2457 GCATGTAACCCAGATGATGATGATGCCATCAAGGTGTGCTGCTCAAGGCCCGGAC 2516  
 Qy 741 PheTyrGlyAsnleuProGlyMetAsnArgProValMetAlaSerLysAspLysleu 760  
 Db 2517 TTTCTGGGTAACTTACCCGGATGATGAACAGGCCCGGTGATGCTTCAAGAAATGATGATGCTC 2576  
 Qy 761 GluIleuGluAspCysleuGluPheSerArgThrAlaLysleuLysValGlnThrIle 780  
 Db 2577 GAGCTGACGAGCTCGCTGAGTTCAGTAGGACAGCAAAAGTTAAAGAAAGTGCAGCAATA 2636  
 Qy 781 ThrThrLysSerAsnSerlleArgGlnGlyLysAsnGluLeuPheProValIleMetAsn 800  
 Db 2637 ACCACCAAGTGAAGTCCATCAGACAGGCGCAAAACAGCTTTTCCCTGTGATGAT 2696  
 Qy 801 GlyLysAspAspValleuTyrCysThrGluIleuGluArgIlePheGlyPheProAlaHis 820  
 Db 2697 GGCAAGAGACGAGTCTTGTGGTGAAGCTGAGCTCGAAAGGATCTTCCCTGCTCAC 2756  
 Qy 821 TyrThrAspValSerAsnMetGlyArgGlyAlaArgGlnLysleuGlyLysArgSerTyr 840  
 Db 2757 TACACGGAGCGTGTCCATATGGCGCGCGCCCTCAGAAAGCTGCTGGAGGAGGCTCTGG 2816  
 Qy 841 SerValProValIleArgHisleuPheAlaProleuLysAspTyrPheAlaCysGlu 859  
 Db 2817 AGTGAACCGGTATCAGACACTGTTTCCCTTGAAGGACTCTTGGCTGTGA 2873  
 RESULT 3  
 LOCUS AF151969 4338 bp mRNA linear ROD 08-JUN-2000  
 DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 1  
 (Dnmt3b) mRNA, complete cds.  
 ACCESSION AF151969  
 VERSION AF151969.1 GI:8347117  
 KEYWORDS  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 4338)

AUTHORS Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and Shen, Y.  
 TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing isoforms in mouse embryonic tissue  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 4338)  
 AUTHORS Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and Shen, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular Biology, Institution of Basic Medical Sciences, Chinese Academy of Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005, P.R.China  
 FEATURES  
 source  
 1. 4338  
 /organism="Mus musculus"  
 /strain="KM"  
 /db\_xref="taxon:10090"  
 /dev\_stage="8-9 day old embryo"  
 1. 4338  
 /gene="Dnmt3b"  
 412..2991  
 /gene="Dnmt3b"  
 /function="methylates cytosine in DNA"  
 /note="alternatively spliced"  
 /codon\_start=1  
 /product="DNA cytosine-specific methyltransferase isoform 1"  
 /protein\_id="AAAF74515.1"  
 /db\_xref="GI:8347118"  
 /translation="MKGDSRHLENEGASGEBCIIVNGNFSQSSDTPDAPSPVLE  
 AITCEPVCPTETGRSSRLSKREVSLLANTQDMDGDRDVEDVDGSDIMPK  
 LITEEDOTRSESPAVTRHSNGTSLERQASPIITRGQRHHVOEYPVFPATR  
 SRRRASSTSPSSPASVDPEMEVTPSVSDVSDGQEQEMDTTOVAESIY  
 GSTEYDDKREKIGDVLWGKIKGSPAMAVMSKATLRQAMRWQWGDGKF  
 SEISADKVALGIFSOHFLATFNKLVSTRKAMTYRLEKARVAGKTFSSPESLED  
 QLEPMLEMAHGFPTGILEKPKKQPVNKSIVRSRNLPRRERKSRRTTN  
 DSAASEPPPKRLKTNVSGKDEDEBSERNAASVTNNKLEDRCLSCGKNVPS  
 FHLFEFGGLCOSGRDFFLEFYVYDGDGYSTVCCERELLCNTSCRCFVEEC  
 LEVLVGAAGAEAKLQEPMSCYVCLPQRCGVLRKRRKDMNMRLODFPTPDLEEP  
 PKLYPAIPAKRRPRLVLSLFDGIATGYLVLELGIKVKYIASVCAESIAYGVKH  
 EGOIKYVNDRKITTKNIEBGFDFYIGSPENDSNVNPARKGIYEGTRLPFEFY  
 HLNLYRPRGDMRPFPFMFENVVAKVNDKDOI SPLCPNPMIDAIKVSANRKY  
 FWNGLPGNRPVMAKNDKLELDCLFERTALKKVQITTSKSDIIRQKDLFPVY  
 MNCKDVLVCTELERIFGPPAHYTDVSNMGRARQLGRSWSVPIRHLFADLVDF  
 ACE"  
 BASE COUNT 1073 a 1124 c 1173 g 968 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,98e-248 Length: 4338  
 Score: 4573.00 Matches: 857  
 Percent Similarity: 99.77% Conservative: 0  
 Best Local Similarity: 99.77% Mismatches: 2  
 Query Match: 99.63% Indels: 0  
 DB: 10 Gaps: 0  
 US-09-720-086-6 (1-859) x AF151969 (1-4338)  
 Qy 1 MetLysGlyAspSerArgHisleuAsnGluGluGlyAlaSerGlyTyrGluGluCys 20  
 Db 412 ATGAAGGAGACACACACATCTGAATGAAGAGAGGTGCACGGGTATGAGAGATGC 471  
 Qy 21 llelleValanGlyAsnPheserArgGlnSerSerAspThrLysAspAlaProSerPro 40  
 Db 472 ATTATCGTTAATGGGAATCTCAGTACCAAGTCTTCAACAGAAAGATGCTCCCTCACCC 531  
 Qy 41 ProValleuGluAlaIleCysThrGluProValCysThrProGluThrArgGlyArgArg 60  
 Db 532 CCACTCTTGGAGGCAATCTGACAGAGCCAGTCTGCACACAGAGACAGAGGCGGACAG 591  
 Qy 61 SerSerSerArgleuSerLysArgGluValSerSerleuLeuAsnTyrThrGlnAspMet 80

592 TCAAGCTCCCGCTGTCTAAGAGGAGGAGTCTCCAGCCTTCTGAATTACACGACGAGCATG 651  
Db  
81 ThrGlyAspGlyAspAlaGlyAspGluValAspAspGlyAsnGlySerAspIleLeuMet 100  
Qy  
652 ACAGGAGATGGAGACAGATGATGAAGTAGATGATGGAAATGGCTCTGATATTCATAATG 711  
Db  
101 ProLeuLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaValArg 120  
Qy  
712 CCAAGCTCAACCCGTGAGACCAAGGACACAGGACGCGCTCTGAAGCCCGCTGTCCGA 771  
Db  
121 ThrArgHisSerAsnGlyThrSerSerLeuGluValArgGlnArgAlaSerProArgIleThr 140  
Qy  
772 ACCGACATAGCAATGGACCTCCAGCTTGGAGAGCAAGAGGCTCCCCAGAAATCAC 831  
Db  
141 ArgGlyArgGlnGlyArgHisValGlnGlyProValGluPheProAlaThrArg 160  
Qy  
832 CGAGGTCCGACGGGCGCCACCATGTGCAAGGATACCCCTGTGGAGTTTCCGGCTTACGAG 891  
Db  
161 SerArgAlaGArgAlaSerSerAlaSerThrProThrProThrProAlaSerVal 180  
Qy  
892 TCTCGGAGACGTCGAGCATCTCTTCAGCAAGCAGCGCATGGTCATCCCTGCGCAGCGTC 951  
Db  
181 AspPheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGln 200  
Qy  
952 GACTTCATGGAAGAAAGTCACACCTAAGAGCGTCAGTACCCCATCAGTTGACTTGGAGCCAG 1011  
Db  
201 AspGlyAspGlnGluMetAspThrThrGlnValAspAlaGluSerArgAspGlyAsp 220  
Qy  
1012 GATGGAGATCAGAGGGTATCGTATACACACAGGTGGATGCGAGAGCATATATGGAGAC 1071  
Db  
221 SerThrGluThrGlnAspAspLysGluPheGlyIleGlyAspLeuValThrGlyLysIle 240  
Qy  
1072 AGCACAGATATCAGAGATATGAAGATTTGGAAATAGGTGACCTCGTGTGGGNAAGATC 1131  
Db  
241 LysGlyPheSerThrProAlaMetValSerThrProLysAlaThrSerLysArgGln 260  
Qy  
1132 AAGGGCTTCTCTGTGTGGCTGSCCATGGTGTGTCTGGAAGCCACCTCCAGCGACAG 1191  
Db  
261 AlaMetProGlyMetArgThrValGlnThrPheGlyAspGlyLysPheSerGluIleSer 280  
Qy  
1192 GCCATCCCGGAATCGCTGGGTACAGTGGTTTGGTGGTGGCAAGTTTCTTGAGATCTCT 1251  
Db  
281 AlaAspLysLeuValAlaLeuGlyLysPheSerGlnHisPheAsnLeuAlaThrPheAsn 300  
Qy  
1252 GCTGCAAACTGGTGGCTCTGGGCTGTTCAGCGCAGCAGCTTTAAATCTGGCTACCTCAAT 1311  
Db  
301 LysLeuValSerThrArgLysAlaMetThrHisThrLeuGluLysAlaArgValArgAla 320  
Qy  
1312 AAGCTGGTTCTTATAGGAGGCCATGTACACACACTCTGGAGAAAGCCAGCGTTTCGAGCT 1371  
Db  
321 GlyLysThrPheSerSerSerProGlyLysLeuGluAspGlnLeuLysProMetLeu 340  
Qy  
1372 GGCAAGACCTTCTCCAGCAGTCTCTGGAGAGTCACTGGAGGACAGCTGAGACCCATCGTG 1431  
Db  
341 GluThrAlaHisGlyGlyPheLysProThrGlyIleGluGlyLeuLysProAsnLysLys 360  
Qy  
1432 GAGTGGGCCACCGTGGCTTCAAGCTACTGGGATCGAGGCTCAAAACCCAAAGAG 1491  
Db  
361 GlnProValValAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProArg 380  
Qy  
1492 CAACCAAGTGGTTAATAAGTCGAAGGTGGCTGCTTCAAGCTAGAGAGTAGAAGCTTAGAACCCAGG 1551  
Db  
381 ArgArgGluAsnLysSerArgArgThrThrAsnAspSerAlaAlaSerGluSerPro 400  
Qy  
1552 AGACCGGAGAAACAAAGTCGAAGACGCAACCAATGACTCTGCTGCTTCTGAGTCCCTCC 1611  
Db  
401 ProProLysArgLeuLysThrAsnSerThrGlyLysAspArgGlyLysAspGluGlu 420  
Qy  
1612 CCACCCAGGCGCTCAAGCAAAATAGCTATGGCGGGAAGACCGAGGGGAGGATGAGGAG 1671  
Db  
421 SerArgGluArgMetAlaSerGluValThrAsnAsnLysGlyValAsnLeuGluAspArgCys 440  
Qy  
1672 AGCCGAGAACGGATGGCTTCTGAGTCTACCAACCAAGGCAATCTCGAAGACCGGCTGT 1731  
Db

441 LeuSerCysGlyLysLysLysAsnProValSerPheHisProLeuPheGluGlyGlyLeuCys 460  
Qy  
1732 TTGTCTCTGGAAAGAAAGAACCTGTGTCTTCCACCCCTCTTTGAGGGTGGGCTCTGT 1791  
Db  
461 GlnSerCysArgAspArgPheLeuGluLeuPheThrMetTyrAspGluAspGlyTyrGln 480  
Qy  
1792 CAGAGTTGGCGGATCGCTTCTCTAGAGCTCTTACATGATATGATGAGAGCGGTATCAG 1851  
Db  
481 SerTyrCysThrValCysCysGluGlyArgGluLeuLeuCysSerAsnThrSerCys 500  
Qy  
1852 TCTTACTGACCGTGTGTGTGAGGCGGTGAACCTGCTGCTGTGCAGTAACACACAGCTGC 1911  
Db  
501 CysArgCysPheCysValGluCysLeuGluValLeuValGlyValGlyThrAlaGluAsp 520  
Qy  
1912 TGCAGATGCTTCTGTGTGGAGTGTCTGGAGGTGCTGTGGCGCAGGCACAGCTGAGGAT 1971  
Db  
521 AlaLysLeuGlnGluProThrSerCysTyrMetCysLeuProGlnArgCysHisGlyVal 540  
Qy  
1972 GCCAAGCTGACGAGAACCTTGGAGCTGTATATGTGCTCTCTCAGCGCTGCCATGGGGTC 2031  
Db  
541 LeuArgArgLysAspThrPheAsnMetArgLeuGlnAspPhePheThrThrAspProAsp 560  
Qy  
2032 CTCCGACGAGGAAGATGGAAACATGCGCTGCAAGACTTCTTCACTACTGATCTTGAC 2091  
Db  
561 LeuGluGluPheGluProProLysLeuTyrProAlaIleProAlaAlaLysArgArgPro 580  
Qy  
2092 CTGGAAGAATTTGAGCCACCCCAAGTTGTACCAGCAATTTCTGCGAGCCAAAGAGAGGCC 2151  
Db  
581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeu 600  
Qy  
2152 ATTAGAGTCTCTCTCTGTTGATGGAATTCACACGCGGTACTTGTGTCTCAAGAGATTG 2211  
Db  
601 GlyIleLysValGluLysTyrIleAlaSerGluValCysAlaGluSerIleAlaValGly 620  
Qy  
2212 GGTATTAAGTGGAAAGTACATGCTCCGAGTCTGTGCAGAGTCCATGCTGTGGGA 2271  
Db  
621 ThrValLysHisGluGlyGlnLysTyrValAsnAspValArgLysIleThrLysLys 640  
Qy  
2272 ACTGTTAAGCATGACAGCCAGATCAATATGTCAATGAGTCCGGAATAATCACCAGAAA 2331  
Db  
641 AsnIleGluGluThrProPheAspLeuValIleGlyGlySerProCysAsnAspLeu 660  
Qy  
2332 AATATTGAAGAGTGGGCGCGCTTCTGACTTGTGATTTGGTGGAGGCCCATGCAATGATCTC 2391  
Db  
661 SerAsnValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPheGlu 680  
Qy  
2392 TCTAACGTCATCTCTCCCGCAAGGTTTATATGAGGCGCACAGGAAGGCTCTTCTTCGAG 2451  
Db  
681 PheTyrHisLeuLeuAsnTyrThrArgProLysGluGlyAspAsnArgProPhePheTrp 700  
Qy  
2452 TTTTACCACCTGCTGAATTTATACCCCGCCCAAGGAGGCGCACACCGTCCATTTCTCTGG 2511  
Db  
701 MetPheGluAsnValValAlaMetLysValAsnAspLysLysAspIleSerArgPheLeu 720  
Qy  
2512 ATGTTCCGAGATGTTGTGGCCATGAAAGTGAATCACAAGAAAGACATCTCAAGATTCTGTG 2571  
Db  
721 AlaCysAsnProValMetIleAspAlaIleLysValSerAlaAlaHisArgAlaArgTyr 740  
Qy  
2572 GCATGTAAACCCAGTATGATGATGCCATCAAGGTGTCTGCTCTCACAGGCGCCGGTAC 2631  
Db  
741 PheThrGlyAsnLeuProGlyMetAsnArgProValMetAlaSerLysAsnAspLysLeu 760  
Qy  
2632 TTCTGGGTGATACCTTACCCGATGAACAGGCCGCTGTATGCTTCAAGAAATGATTAAGTCTC 2691  
Db  
761 GluLeuGlnAspCysLeuGluPheSerArgThrAlaLysLeuLysValGlnThrIle 780  
Qy  
2692 GAGCTGACGAGTCTGCTGGAGTTTCAAGTGGACAGCAAGTTTAAAGAAAGTGCAGACAATA 2751  
Db  
781 ThrThrLysSerAsnSerIleArgGlnGlyLysAsnGlnLeuPheProValValMetAsn 800  
Qy  
2752 ACCCAAGTGCAGACTCTCATCAGAGGCGCAAAACCAGCTTTTCCCTGTGTAGTCAATGAAT 2811  
Db

QY 801 GlytysaparpvalletrtccyethrctgltuenglualrgllepheglypheproalaHis 820  
Db 2612 GGCAGAGCAGCGTTTGTGTGTGCACTGAGCTCCAAAGAACTTCGCGCTTCCCTGCAC 2871  
QY 821 TythrapsvalserasmetglYarglYalaaarglntsluenglYargSetrtp 840  
Db 2872 TACCGACGCTGTCCACATGCGCCGCGCCCGCTCAGAAAGCTGCTGGAGGTCCTGG 2931  
QY 841 SerValProValIleArgHleuPheAlaProleuYsAspTYrPheAlaCysGlu 859  
Db 2932 AGGTACCGGTCATCAGACACCTGTTGCCCTTGAAGAGACTTCTTGCTGTGAA 2988  
RESULT 4  
AF068627 4135 bp mRNA linear ROD 06-DEC-1999  
LOCUS Mus musculus DNA cytosine-5 methyltransferase 3B2 (Dnmt3b) mRNA,  
DEFINITION alternatively spliced, complete cds.  
ACCESSION AF068627  
VERSION AF068627.2 GI:6449471  
KEYWORDS  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
REFERENCE 1 (bases 1 to 4135)  
AUTHORS Okano, M., Xie, S., and Li, E.  
TITLE Cloning and characterization of a family of novel mammalian DNA  
JOURNAL Nat. Genet. 19 (3), 219-220 (1998)  
MEDLINE 98324766  
PUBMED 9662389  
REFERENCE 2 (bases 1 to 4135)  
AUTHORS Xie, S., Okano, M., and Li, E.  
TITLE Direct Submision  
JOURNAL Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA  
REFERENCE 3 (bases 1 to 4135)  
AUTHORS Okano, M., Chijiwa, T., Sasaki, H., and Li, E.  
TITLE Direct Submision  
JOURNAL Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA  
REMARK Sequence update by submitter  
COMMENT On Nov 18, 1999 this sequence version replaced gi:3327981.  
FEATURES  
source location/Qualifiers  
1..4135 /organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/map="84.0 cM"  
1..4135 /gene="Dnmt3b"  
1..4135 /gene="Dnmt3b"  
misc\_feature /note="similar to EST sequences deposited in GenBank  
Accession Numbers AA116694, AA119979, AA177277, AA210568,  
AA407106, and AA575617"  
269..2788 /gene="Dnmt3b"  
/function="de novo DNA methylation"  
/note="alternatively spliced product, contains Cys-rich  
region; C-terminal region is similar to corresponding  
region of DNA cytosine-5 methyltransferase 3A"  
/codon\_start=1  
/product="DNA cytosine-5 methyltransferase 3B2"  
/protein\_id="AAC40179.2"  
/db\_xref="GI:6449472"  
/translation="MKGDSRLNLEEBEGASYEEICIVNGNFSQSSDPTKDAPEPVLE  
AICTPEYCTETRGSSRLSKREVSILNTYQDMGDRDDEVDDGSDILMPK  
LTREKDTRTRESBPAYRTSHNSGTSLEORASPRITRGQRHNVQEVPEFAPR  
SRBRASSASTSPASVDFMEVTPKVSVPVLTSDGQEGMDTTOVAESD  
GDSREYDDKEFGGLVWKIKFGFWMPAMVYSKRTSKROMPKRWOMQGDGCF  
SEISADLVAGLFSQHFNLATNTKLVSYRKAMHTLEKRVRAKTFSSPESLSD  
QLKFMLEWAGHGFPTGIEGLKPKKQPKENKSRRTTNDASASPPKRLKTNYSYG

BASE COUNT 1043 a 1048 c 1106 g 938 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 9,09e-243 Length: 4135  
Score: 4472.00 Matches: 839  
Percent Similarity: 97.67% Conservative: 0  
Best Local Similarity: 97.67% Mismatches: 0  
Query Match: 97.43% Indels: 20  
DB: 10 Gaps: 1  
US-09-720-086-6 (1-859) x AF068627 (1-4135)  
QY 1 MetLygIyAspSerArghIslEaAsnglUglUglYalaseGlyTYrGluGluCys 20  
Db 269 ATGAAGGAGACACACACATCTGATGAAGAAGAGGCTCCAGCGGATGAGAGAGTGC 328  
QY 21 IleIleValAsngIyaAspSerAspGlnSerSerAspThryAspAlaProSerPro 40  
Db 329 ATTATCGTTATGGAACTTCAGTACCACTCTCCAGACACGAAAGATGCTCCCTCACCC 388  
QY 41 ProValleuGlualalIeCysThrGluProValCysThrProGluThraGluYarg 60  
Db 389 CCACTCTTGAGGCAACTCTCAGACAGCCAGCTCTGCACACAGAGCCAGAGCCGACG 448  
QY 61 SerSerSerArpLeuSerLysArgGluValSerSerLeuAsnTYrThrglnAspMet 80  
Db 449 TCAAGCTCCGGCTGTGTTAAGAGGAGGCTCCAGCTTGTGAATTACAGCAGAGCATG 508  
QY 81 ThrGlyAspGlyAspArgAspGluValAspAspGlyAsngIySerAspIleLeuMet 100  
Db 509 ACAGAGATGAGACACAGATGATGAAGTATGATGATGATGATGATGATGATGATGATG 568  
QY 101 ProLyseuThraArgIuThrLysAspThraArgThraArgSerGluSerProAlaValArg 120  
Db 569 CCAAGCTCACCCGTCGAGACCAAGAGACAGACAGCGGCTCTGAAGCCCGGCTGTCGA 628  
QY 121 ThraArgIserAsngIyThrrSerSerLeuGluArgIlnArgIlnAspProArgIleThr 140  
Db 629 ACCGACATAGCAATGGACCTCCAGCTTGAGAGGCAAGAGGCTCCCCAGATACCC 688  
QY 141 ArgGlyArgGlnGlyArgHshIvalGlnGluTYrProValGluPheProAlaThraArg 160  
Db 689 CGAGTCGGCAGGCGCCGCCACCATGTCCAGAGTACCTTGAGATTCCGGCTCACAGG 748  
QY 161 SerArgHArgArgAlaSerSerSerAlaSerThrProTrpSerSerProAlaSerVal 180  
Db 749 TCTCGAGAGCTCGAGCATGCTCTTGCAGAACGACGCAATGTCATCCCTGCAGAGCTC 808  
QY 181 AspPheMetGluGluValThrProLyseValSerThrProSerValAspLeuSerGln 200  
Db 809 GACTTCATGGAAGAAGAGACCTTAAAGGCTCAGTACCCCATCAGTTGACTTGAGAGCAG 868  
QY 201 AspGlyAspGlnGluGlyMetAspThrThrglnValAspAlaGluSerArgAspGlyAsp 220  
Db 869 GATGAGATACGAGAGGTATAGATACCAACAGGTGATCAGAGAGCAGAGATGAGAGC 928  
QY 221 SerThrgIuTYrGlnAspAspLygIuPheGlyIleGlyAspLeuValTrpGlyLysIle 240  
Db 929 AGCACAAGATATCAGATGATTAAGAAGTTTGATAGTATGATGATGATGATGATGATG 988  
QY 241 LygGlyIyAspSerTrpTrpProAlaMetValValSerTrpLySalatThrSerLyBatrgln 260  
Db 989 AAGGCTTCTCTCGTGTGCTGCTGCATGAGTGTCTCGAAGAACCACTCAGGCGCAG 1048





```
FEATURES
source
    Location/Qualifiers
        1..4163
            /organism="Mus musculus"
            /strain="RM"
            /db_xref="taxon:10090"
            /dev_stage="8-9 day old embryo"
            1..4163
            /gene="Dmmt3b"
            297..2816
            /gene="Dmmt3b"
            /function="methylates cytosine in DNA"
            /note="alternatively spliced"
            /codon_start=1
            /product="DNA cytosine-specific methyltransferase isoform 6"
            /protein_id="AA074520.1"
            /db_xref="GI:8347131"
            /translation="MKGDSPRLNDEEGASCEECIIVNGNFSQSSDPTKDAPEPVLE
            AITCEPVCTPETGRSSSRSLSKREVSLNITQMDGDDRDDEVDDGSDILMPK
            LIREIDTRTRSSPAVTRTHSNGTSLERQASPIITRQGRHHQVEYVPEPAIR
            SRRRASSASVTPSSPASYDMEBVTPEKSVTPSVLSQDQEGMDITQVASELY
            GDSFEYODDKFPGIDLVWGKIKGFSWPMAMVSMKATSKROMPGMRVOMQMGDF
            SEISADKVALGIFSOHFNLATFNKLVSRYKAMVTHLEKARVAGTRPSSPESIED
            QLRMLERWAGSRFKPTIGLAKPNKKQPEKSKRRPTTNDASASEPPPKLKTNSIG
            KQGBDESEKERVASEVTNNKGNLEDRCLSGKKNPVSPFLEGGLQSGCRDRIEL
            FMYDGDGQSYCTVCCEGRELICNSGCCRCCEVLVAGAGTAEBAKQIEPWS
            CYMCLPQRCGVARRKDMNMLQDEFTTDPDLEEPFLYAIPAKRRPIRVLSL
            FDIATGVLVLEKELIKVEKYIASEVASEIIVGTVYHGOIYVNDVRKITTKNIE
            WGPBPLVIGSSPCNDLSNVPARKLYEGTGRI.PPEFYHILANTREKGNRPFPMF
            ENVVAMKANDKDISRFACNPVIMDIKISAHARARYFMGNLPGMKRPMASKNDL
            ELDDCLEFSTALVKVOTITTKSNSIRQKNLFPVMNGKDDVIMCTELERIFGFP
            AHYTDVSNMGRKORLQGSMSVPVIRHLPFKDYFACE"
BASE COUNT      1032 a      1074 c      1128 g      929 t
ORIGIN
Alignment Scores:
Pred. No.:      8.3e-242      Length:      4163
Score:          4455.00      Matches:      837
Percent Similarity: 97.44%      Conservative: 0
Best Local Similarity: 97.44%      Mismatches: 2
Query Match:    97.06%      Indels:      20
DB:             10      Gaps:        1
US-09-720-086-6 (1-859) x AF151974 (1-4163)
Q      1      MetLysGlyAspSerArgHisLeuAsnGluGluGluAlaSerGlyTyrGluGluCys 20
Db     297      ATGAAAGGAGACACACACATCTGAATGAAGAAGAGGTGCACGGGTATAGAGAGAGGC 356
Qy     21      ILeIleValAsnGlyAsnPhseSerAspGlnSerSerAspThrLysAspAlaProSerPro 40
Db     357      ATTATCTTATAGGAACTTCAGTACCAAGTCCCTCAGACACAGAAAGATGCTCCCTCACCC 416
Qy     41      ProValLeuGluAlaIleCysThrGluProValLysThrProGluThrArgGlyValArg 60
Db     417      CCAAGTCTTGAGGAGAACTGCACAGACCAAGTGCACACAGAGACACAGAGCCGACAG 476
Qy     61      SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTyrThrGluAspMet 80
Db     477      TCAAGCTCCCGAGCTGTCTAAGAGGAGAGGTCTCCAGCTTCTGAATTCACGCAAGACATG 536
Qy     81      ThrGlyAspGlyAspArgAspGluValAspAspGlyAsnGlySerAspIleLeuMet 100
Db     537      ACAAGAAATGGAGACAAAGATGATGAAGTAGATGATGGAAATGCTCTGATATTTCTAATG 596
Qy     101      ProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaValArg 120
Db     597      CCAAAAGCTCAACCCCTGAGACCAAGACACCGAGCGCTCTGAAGAAGCCCGCTGTCCGA 656
Qy     121      ThrArgHisSerAsnGlyThrSerSerLeuGluArgGluArgAlaSerProArgIleThr 140
Db     657      ACCGGAATGCAATGGAGACTCCAGCTTGAGAGGACAAAGAGCTCCCGCCCAAGATACCC 716
```

```
Qy     141      ArgGlyValArgGlnGlyArgHisIleValGlnGluThrProValGluPheProAlaThrArg 160
Db     717      CGAGGTGGGAGAGGCCGCCACCATGTGCAGAGTACCTGTGAGATTCCGGCTTACAGAG 776
Qy     161      SerArgArgArgArgAlaSerSerSerSerAlaSerThrProThrSerSerProAlaSerVal 180
Db     777      TCTCGAGAGAGTCCAGAGATGTCTTCAGCAAGACGCGCATGTGATCATCCCTGCCAGGTC 836
Qy     181      AspPheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGln 200
Db     837      GACTTCATGGAAGAAGAGACACCTTAAGAGGCTCAGTACCCCATAGTGAATCTGAGCCAG 896
Qy     201      AspGlyAspGlnGluGlyMetAspThrThrGlnValAspAlaGluSerArgAspGlyAsp 220
Db     897      GATGAGATCAGAGAGGAGTATGATACCAACAGAGTGATCAGAGACATATATAGAGAC 956
Qy     221      SerThrGluThrGluAspAspLysGluPheGlyIleIleLysAspLeuValTTPGlyLysIle 240
Db     957      AGCAGAGATATCAGATGATTAAGAGATTGGAAATAGGTACCTCGATGTGGGAAAGATC 1016
Qy     241      LysGlyPheSerTPTTPProAlaMetValSerTrrPlyAsaIaThrSerLysArgGln 260
Db     1017      AAGGCTTCTCTGTGTGCTGCTCCATGTGTGTCTCTCGAAAGCCACTCCAAAGCAGAG 1076
Qy     261      AlaMetProGlyMetArgTrrPValGlnTrrPheGlyAspGlyLysPheSerGluIleSer 280
Db     1077      GCCATGCCCGGAATGCCCTGGGTACAGTGGTTGGTATGGCAAGTTTCTGATGATCTCT 1136
Qy     281      AlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsn 300
Db     1137      GCTGACAAACTGGGGCTCGGGCTCTTCAGCCAGACACTTAACTCGGCATCTTCAAT 1196
Qy     301      LysLeuValSerTrrArgLysAlaMetTrrHisThrLeuGluLysValIaArgValAla 320
Db     1197      AAGCTGTTCTTATAGAAAGCCCATATACACACTCTGGAGAAAGCCAGGGTTGAGCT 1256
Qy     321      GlyLysThrPheSerSerSerProGlyLysSerLeuGluAspGlnLeuLysProMetLeu 340
Db     1257      GGCAAGACCTTCTCCACAGAGCTCGGAGAGTCACTGGAGGACCAAGCTGAAGCCATGCTG 1316
Qy     341      GluTrrPalaHisGlyGlyPheLysProThrGlyIleGluGlyLeuLysProAsnLysLys 360
Db     1317      GAGTGGGCCCAAGCTGCTTCAAGCTTACGAGTCCAGAGGCTCAAAACCAACAAAGAG 1376
Qy     361      GluProValValAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProArg 380
Db     1377      CAACCC----- 1382
Qy     381      ArgArgLysAsnLysSerArgArgArgThrThrAsnAspSerAlaIaSerGluSerPro 400
Db     1383      -----GAGAAACAAGTCCAGAAAGCGACAAACCAATCACTGTGCTTCTGATGCCCCC 1436
Qy     401      ProProLysArgLeuLysThrAsnSerTyrGlyLysAspArgGlyGluAspGluGlu 420
Db     1437      CCAACCCAGGCGCTCAAGAAATAGCTATGCGGGAAGACCAAGGAGGAGAGAGAGAG 1496
Qy     421      SerArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCys 440
Db     1497      AGCCGAGAAAGGATGGTCTTGAAGTCAACCAACAAAGGCAATCTGGAAAGACGGCTGT 1556
Qy     441      LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyLysCys 460
Db     1557      TTGTCTGTGTGAAAGAAAGAACCTGTGTCTCTTCAACCCCTCTTTGAGGGTGGCTCTGT 1616
Qy     461      GlnSerCysArgAspArgPheLeuGluLeuPheTyrMetTyrAspGluAspGlyTyrGln 480
Db     1617      CAGAGTTGCCGGAGTCCCTCTTCAAGACTCTTCTAATGATGATGAGAGACGGCTATCAG 1676
Qy     481      SerTrrCysThrValCysCysGluGluArgGluLeuLeuLeuLeuSerAsnThrSerCys 500
Db     1677      TCTCACTGCAACCGTGTGTGTGAGGGCGGTGAATGTGTGTGTGACATTAACAAAGCTGC 1736
Qy     501      CysArgCysPheCysValGluCysLeuGluValLeuValGlyIaGlyThrAlaGluAsp 520
```

Db 1737 TGCAGATGCTTCTGTGTGGAGTGTCTGGAGGTGCTGTGGGGCGCAGGCACAGCTGAGGAT 1796  
Qy 521 AlaLysLeuGlnGluProTyrPserCysTyrMetCysLeuProGlnArgCysHisGlyVal 540  
Db 1797 GCCAAGCTGCAGAACCTGGAGCTGCTATATGTGCTCCCTCAGCGCTGCCATGGGGTC 1856  
Qy 541 LeuArgTyrGlnGluPheAspTyrPheMetCysLeuGlnGluPhePheThrThrAspProAsp 560  
Db 1857 CTCGACGCGAGAAAGATTGGAACATCGCGCTGCAAGACTTCTTCACTACTGATCTCTGAC 1916  
Qy 561 LeuGlnGluPheGluProProGlyLeuTyrProAlaIleProAlaAlaLysArgCysPro 580  
Db 1917 CTGGAAGAATTGACCCACCAAGTTGTACCCAGCAATTCCTGCAGCCAAAGAGGCC 1976  
Qy 581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeu 600  
Db 1977 ATTAGAGTCTCTCTCTTGTATGGAATTGCAACGGGGTACTTGGTGCTCAAGGAGTTG 2036  
Qy 601 GlyIleValGluLysTyrIleAlaSerGluValCysAlaGluSerIleAlaValGly 620  
Db 2037 GGTATTAAAGTGGAAAGTACATTGCTCCGAAAGTCTGTGCAGAGTCCATCGCTGGGA 2096  
Qy 621 ThrValLysHisGluGlyGlnIleLysTyrValAsnAspValArgLysIleThrLysLys 640  
Db 2097 ACTGTTAAGCATGAAGGCCAGATCAATATGTCAATGACGTCCGGAATAATCACCAGAAA 2156  
Qy 641 AsnIleGluGluTyrGlyProPheAspLeuValIleGlyGlySerProCysAsnAspLeu 660  
Db 2157 AATATTGAAGAGTGGGGCCCGTTCGACTTGGTGTGGAAGCCCATGCAATGATCTC 2216  
Qy 661 SerAsnValAspProAlaArgGlyLeuTyrGluGlyThrGlyArgLeuPhePheGlu 680  
Db 2217 TCTACGTCATCTCTGCCCGCAAGGTTTATATGAGGGCACAGGAAGCTCTTCTTCGAG 2276  
Qy 681 PheTyrHisLeuLeuAsnTyrThrArgProLysGluGlyValAsnArgProPhePheTyr 700  
Db 2277 TTTTACCTCTCTGAATATATACCCGCCCAAGAGGGCGACACCGTCCATCTTCTCG 2336  
Qy 701 MetPheGluAsnValAlaMetLysValAsnAspLysLysAspIleSerArgPheLeu 720  
Db 2337 ATGTTTCGAAATGTTGTGGCCATGAAAGTGAATGACAAAGAAAGACATCTCAAGATTCCTG 2396  
Qy 721 AlaCysAsnProValMetIleAspAlaIleLysValSerAlaAlaHisArgAlaArgTyr 740  
Db 2397 GCATGTAACCCAGTCATGATCGATCGCATCAAGGTGTCTGCTGCTCACAGGGCCCGGTAC 2456  
Qy 741 PheTyrGlyAsnLeuProGlyMetAsnArgProValMetAlaSerLysAsnAspLysLeu 760  
Db 2457 TTCTGGGGTAACCTTACCCGGAATGAACAGGCCCGTGTGATGGCTTCAAGAAATGATAAGCTC 2516  
Qy 761 GluLeuGlnAspCysLeuGluPheSerArgThrAlaLysLeuLysValGlnThrIle 780  
Db 2517 GAGCTGCAGGACTGCTCGAGTTTCAGTAGCAGCAGCAAGTTAAAGAAAGTCACAGCAATA 2576  
Qy 781 ThrThrLysSerAsnSerIleArgGlnGlyAsnGlnLeuPheProValValMetAsn 800  
Db 2577 ACCACCAAGTCAGATCTCATCAGACAGGCAAAACACAGCTTTTCCCTGTGATCATGAT 2636  
Qy 801 GlyLysAspAspValLeuTyrCysThrGluLeuGluArgIlePheGlyPheProAlaHis 820  
Db 2637 GGCAAGGACGAGCTTTTGTGTGTCACGTGACGTGCAAGGATCTTCGGTTCCTCCCTGCAC 2696  
Qy 821 TyrThrAspValSerAsnMetGlyArgGlyAlaArgGlnLysLeuGlyArgSerTyr 840  
Db 2697 TACACGAGCTGTCAACATGGCGCGCGCCGTCAGAAAGCTCTCGGAGAGTCTCGG 2756  
Qy 841 SerValProValIleArgHisLeuPheAlaProLeuLysAspTyrPheAlaCysGlu 859  
Db 2757 AGTGTACCGGTATCAGACACCTGTTTGGCCCTTGAAGGACTACTTGGCTGTGAA 2813  
RESULT 6  
AF151970

LOCUS AF151970 4278 bp mRNA linear ROD 08-JUN-2000  
DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 2  
(Dnmt3b) mRNA, complete cds.  
ACCESSION AF151970  
VERSION AF151970.1 GI:8347119  
KEYWORDS  
SOURCE Mus musculus.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 4278)  
AUTHORS Yin, B., Chen, Y.T., Zhu, M., Luo, Y.J., Zhu, N., Xu, S.C., Wu, G.Y. and  
Shen, Y.  
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing  
isoforms in mouse embryonic tissue  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4278)  
AUTHORS Yin, B., Chen, Y.T., Zhu, M., Luo, Y.J., Zhu, N., Xu, S.C., Wu, G.Y. and  
Shen, Y.  
DIRECT SUBMISSION  
SUBMITTED (17-MAY-1999) Department of Biochemistry and Molecular  
Biology, Institution of Basic Medical Sciences, Chinese Academy of  
Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,  
P.R.China  
FEATURES  
source  
1..4278  
/organism="Mus musculus"  
/strain="KM"  
/db\_xref="taxon:10090"  
/dev\_stage="8-9 day old embryo"  
gene 1..4278  
/gene="Dnmt3b"  
CDS 412..2931  
/gene="Dnmt3b"  
/function="methylates cytosine in DNA"  
/note="alternatively spliced"  
/codon\_start=1  
/product="DNA cytosine-specific methyltransferase isoform 2"  
/protein\_id="AAF74516.1"  
/db\_xref="GI:8347120"  
/translation="MKGDSRLNEEGAGSYECIIIVNGNFSDQSDTKDAPSPVLE  
LIRETKDTRTRESPAVRTRHSNGTSLERQASPRITRGRGRHHVQVEYVEFPAT  
SRRRASSTPSSPASVDPMEEVTPKSVSTPSVDLSQDQDMOTQVDAES  
GDSLEYDDKEFGIDLVGKIGFVSWPMAMVSWKATSKRAMPCMRMVMQVFGDGR  
SEISDKLVALGLFSQFNLATFNKLYSRKAMHYHTLEKARVRAGKTFFSSFGSDGL  
QLKPMLEWAHGGFKPTGIEGLKNKQPNKSRRTTNDSSAESPFPKRLKTNISYGG  
KRGDEESERMASEVTNNKGNLDRCLSCGKNPVSFHLPEGGLCQSCDRFLLEL  
FYWDDGYQSYCTCCGRELILCSNCCSCFCVECLEVLVGAGTAADAKLPVLSL  
CYMLPQRCHGVLRRRKDNMLDPTTDLLEBEPKLPYPAIAPAKRPIKRVLSL  
FDGIATGVLVLELGIKVEKVIASEVCAESIAVGVKHEGQIKYVNDVRKTKNIRE  
WGPFDLVIGGSPCDNLNPNPARKGLYEGTGRLPFFHYHLNATYRPEKGNRPFFMFP  
ENVAVKNDKKDISRFLACNPVMIDAIKVSAAHARAYPMGNLPGMNRPMVASKNDKL  
ELQDCLEFSRTAKLKKVQITTKTNSIRQSGKNLPVVMNGKDDVLWCTELERIFGFP  
AHYTDVSNMGRGAKOKLLGRSWSVPVIRHLFAPLKDYPACE"  
BASE COUNT 1055 a 1113 c 1154 g 956 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 8,58e-242 Length: 4278  
Score: 4455.00 Matches: 837  
Percent Similarity: 97.44% Conservative: 0  
Best Local Similarity: 97.44% Mismatches: 2  
Query Match: 97.06% Indels: 20  
DB: 10 Gaps: 1  
US-09-720-086-6 (1-859) x AF151970 (1-4278)  
Qy 1 MetLysGlyAspSerArgHisLeuAsnGluGluGluGlyAlaSerGlyTyrGluGluCys 20  
Db 412 ATGAAGGAGAGACAGACATCTTGAATGAAGAAGAGGGTCCAGCGGTATGAGGAGTGC 471

QY 21 IleIleValaAsnGlyAsnPheSerAspGlnSerSerAspThrIlyAspAlaProSerPro 40  
 Db 472 ATATTCGTTAAAGGAACTTCAAGTACCACTTCAGACAGAGAAAGTACCTCCACCC 531  
 QY 41 ProValLeuGluAlaIleCysThrGluProValCysThrProGluThrArgGlyArgArg 60  
 Db 532 CCACTCTTGGAGGCAATTCGACAGAGCCACTCTGCACACCAAGAGCCAGAGCCGAGG 591  
 QY 61 SerSerSerArgLeuSerIlyArgGluValSerSerLeuLeuAsnIlyThrGlnAspMet 80  
 Db 592 TCAAGCTCCCGGCTGTCTTAAAGAGGAGGTCTCAAGCTTCTGAATTACACGACAGCATG 651  
 QY 81 ThrGluValAspGlyAspArgAspAspGluValAspAspGlyAsnGlySerAspIleLeuMet 100  
 Db 652 ACAGAGAAATGAGACAGAGATGAAATGATATGATGGAATGGCTCTGAATATCTAATG 711  
 QY 101 ProIlyLeuThrArgGluThrIlyAspThrArgThrArgSerGlnSerProAlaValArg 120  
 Db 712 CCAAGCTCACCCCTGAGACCAAGGACCCAGAGCCCTCTGAAGCCCGGCTGTCCGA 771  
 QY 121 ThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArgAlaSerProArgIleThr 140  
 Db 772 ACCGACATAGCATAGGAGACCTCCAGCTTGGAGAGGAAAGCTCCCCAGAAATCAC 831  
 QY 141 ArgGluArgGlnGlyArgHisValGlnGluIlyProValGluPheProAlaThrArg 160  
 Db 832 CGAGGTGGGACGGGCGCCACCATGTGAGAGTACCTGTGGAGTTTCCGCTACACAG 891  
 QY 161 SerArgArgArgArgAlaSerSerSerAlaSerThrProTrpSerSerProAlaSerAl 180  
 Db 892 TCTGGAGACGTCAGACATCGTCTTACAGACAGCCCATGCTCATCCCTCCAGACCTC 951  
 QY 181 AspPheMetGluGluValThrProIlySerValSerThrProSerValAspLeuSerGln 200  
 Db 952 GACTTCATGGAAGAAGTACACCTTAAGAGCGTCACTACCATCAGTTGACTTGAGCCAG 1011  
 QY 201 AspGluAspGlnGluGlyMetAspThrThrGlnValAspAlaGluSerArgAspGlyAsp 220  
 Db 1012 GATGAGATCGAGAGGATGATGATACCAACAGAGTGTGAGAGACATATATGAGAAC 1071  
 QY 221 SerThrGluIlyArgIlyAspAspIlyGluPheGlyIleGlyAspLeuValThrGlyIlySerIle 240  
 Db 1072 AGCAGAGATCGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1131  
 QY 241 IlySerIlyPheSerIlyTrpProAlaMetValIleSerTrpIlyAlaThrSerIlyArgGln 260  
 Db 1132 AAGGCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1191  
 QY 261 AlaMetProGlyMetArgTrpValGlnTrpPheGlyAspGlyIlyPheSerGluIleSer 280  
 Db 1192 GCCATGCCCCGGAATGCGCTGGGTACAGGTGATGATGATGATGATGATGATGATGATG 1251  
 QY 281 AlaAspIlySerValAlaIleGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsn 300  
 Db 1252 GCTACAAACCTGGGCTCTGGGCTGTTCAAGCCAGCACTTAACTGGCTACCTTCAT 1311  
 QY 301 IlyLeuValSerIlyArgIlyAlaMetIlyHisThrLeuGluIlyValAlaArgValArgAla 320  
 Db 1312 AAGCTGCTTCTTATAGGAAGCCATGATACACACTCTGGAGAAAGCCAGGCTTCAGACT 1371  
 QY 321 GlyIlyThrPheSerSerSerProGlyIlySerLeuGluAspGlnLeuIlyProMetLeu 340  
 Db 1372 GGCAGAGACCTTCTCTGAGAGTCTGAGAGTCACTGAGAGCCAGACTGAACCCCATGCTG 1431  
 QY 341 GluTrpAlaHisGlyIlyGlyPheIlyProThrGlyIleGlyIlyLeuIlyProAsnIlyIly 360  
 Db 1432 GAGTGGGCCCCAGGGGCTTCAAGCTTCAAGTCAAGGAGCTTCAAAACCAAGAGAG 1491  
 QY 361 GlnProValIlyAsnIlySerIlyValIlyArgSerAspSerArgAsnLeuGluProArg 380  
 Db 1492 CAACA----- 1497  
 QY 381 ArgArgGluAsnIlySerArgArgArgThrThrAsnAspSerAlaIleSerGluSerPro 400

Db 1498 -----GAGAACAAAGTCAAGACGACCAATGACTCTGCTTCTGAGTCCCC 1551  
 QY 401 ProProIlyAspLeuIlyThrAsnSerIlyGlyIlyAspAspArgGlyIlyAspGluGlu 420  
 Db 1552 CCACCCAAAGCCCTTCAAGACAAATAGCTATGGCGGAAAGACCCAGGGAGAGATGAGAG 1611  
 QY 421 SerArgIlyArgMetAlaSerGluValThrAsnAsnIlySerGlyAsnLeuGluAspArgCys 440  
 Db 1612 AGCCGAGAACGATGGCTTCTGAAGTCAACCAAGAGGCAATCTGGAAGACCGCTGT 1671  
 QY 441 LeuSerCysGlyIlyIlyAsnProValSerPheHisProLeuPheGluGlyIlyLeuCys 460  
 Db 1672 TTGTCTGTGAAAGAAAGAACCTGTGTCTTCAACCCCTCTTGGAGGGGCTCTGT 1731  
 QY 461 GlnSerCysArgAspArgPheLeuGluLeuPheIlyMetIlyAspGluAspGlyIlyGln 480  
 Db 1732 CAGAGTCCGGGATCGCTTCTTGAAGCTCTTCACTATGATGAGAGCGCTATAG 1791  
 QY 481 SerIlyCysThrValCysCysGluGlyArgGluLeuLeuLeuCysSerAsnThrSerCys 500  
 Db 1792 TCTTACTGACACGTGTGCTGTGAGGCGGTGAATGCTGTGCTGTGCAATACAAAGCTGC 1851  
 QY 501 CysArgCysPheCysValGluCysLeuGluValLeuValGlyIlyThrAlaGluAsp 520  
 Db 1852 TGCAATGCTTCTGTGTGAGATGTGTGAGGTGTGTGTGGCGGAGGACAGCTGAGGAT 1911  
 QY 521 AlaIlyLeuGlnGluProTrpSerCysIlyMetCysLeuProGlnArgCysHisGlyVal 540  
 Db 1912 GCCAAGCTGACGGAACCTGAGAGCTGCTATATGTGCTCTCCCTACGCGCTGCAATGGGATC 1971  
 QY 541 LeuArgArgArgIlyAspTrpAsnMetArgLeuGlnAspPhePheThrThrAspProAsp 560  
 Db 1972 CTCCGACGACAGAAAGATTGGAACATGGCGCTGCAAGACTTCTTCACTAGATCTTAC 2031  
 QY 561 LeuGluGluPheGluProProIlyLeuIlyProAlaIleProAlaAlaIlyArgArgPro 580  
 Db 2032 CTGGAAGATTGAGCCACCCAGCTGTATGCCCAATCTCTGACGACCAAGAGGCCC 2091  
 QY 581 IleArgValIleuSerLeuPheAspGlyIleAlaThrGlyIlyLeuValIleuLeuGluLeu 600  
 Db 2092 ATTAGACTCTGTCTCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2151  
 QY 601 GlyIlyLeuValGluIlyIlyIleAlaSerGluValCysAlaGluSerIleAlaValGly 620  
 Db 2152 GGTATTAAAGTGAAGAAAGTATGATGCTCCGAACTGTGTGAGATGATGATGATGATGATG 2211  
 QY 621 ThrValIlyHisGlyIlyGlnIleIlyIlyValIlyAsnAspValIlyArgIlyIleThrIlyIly 640  
 Db 2212 ACTGTTAAGCATGAAAGCCAGATCAAAATATGTCAATGACGTCCGGAATAATCACCAAGAAA 2271  
 QY 641 AsnIleGluGluTrpGlyProPheAspLeuValIleGlyIlySerProCysAsnAspLeu 660  
 Db 2272 AATATTGAAGTGGGGCGCTGTGACTGTGTGATGATGATGATGATGATGATGATGATGATG 2331  
 QY 661 SerAsnValAsnProAlaArgIlyGlyLeuIlyTrpGlyIlyThrGlyArgLeuPhePheGlu 680  
 Db 2332 TCTAAGTCATCTGCTGCGGAAAGGTTTATAGAGGCAAGAGAGCTTCTTTCAG 2391  
 QY 681 PheIlyHisLeuLeuAsnIlyThrArgProIlyGluGluIlyAspAsnArgProPhePheTrp 700  
 Db 2392 TTTTACCACTTGTGAATTATACCGCCCAAGAGAGGCGCAACCGCTCATTTCTTCTGG 2451  
 QY 701 MetPheGluAsnValIlyAlaMetIlyValIlyAsnAspIlyIlyAspIleSerArgPheLeu 720  
 Db 2452 ATGTTGAGAAATGTTGGCCATGAAAGTGAATGACAAAGAAACATCTCAAGATCTCTG 2511  
 QY 721 AlaCysAsnProValMetIleAspAlaIleIlyValSerAlaAlaHisArgAlaArgIly 740  
 Db 2512 GCATGTAAACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2571  
 QY 741 PheTrpGluAsnLeuProGlyMetAsnArgProValMetAlaSerIlyAsnAspIlyLeu 760

Db 2572 TTCTGGGTAACTACCCGGAATGAACACAGGCCCGGTGATGGCTTCAAAGAATGATAAGCTC 2631  
Qy 761 GLuLeuGlnAspCysLeuGluPheSerArgThrAlaLysLeuLysLeuValGlnThrIle 780  
Db 2632 GAGTCGAGGACTGCTGGAGTTTCAGTAGGACAGCAAAAGTTAAAGAAAGTGCACACAATA 2691  
Qy 781 ThrThrLysSerAsnSerIleArgGlnGlyLysAsnGlnLeuPheProValValMetAsn 800  
Db 2692 ACCACCAAGTCGAATCCATCAGACAGGCGCAAAACACAGCTTTTCCCTGTAGTCATGAAT 2751  
Qy 801 GlyLysAspAsnValLeuTrpCysThrGluLeuGluArgIlePheGlyPheProAlaHis 820  
Db 2752 GGCAGAGACAGCTTTTGTGTGACCTGAGCTCGAAGGATTTTCGGCTTCCCTGCTCAC 2811  
Qy 821 TyrThrAspValSerAsnMetGlyArgGlyAlaArgGlnLysLeuLysLeuGlyArgSerTrp 840  
Db 2812 TACACGGAGCTGTCAACATGGCGCGCGCGCCGTCAGAAGCTCTCTGGGAGGTCCTGG 2871  
Qy 841 SerValProValIleArgHisLeuPheAlaProLeuLysAspTyrPheAlaCysGlu 859  
Db 2872 AGTGTACCGGTATCAGACACCTGTTTGGCCCCCTTGAAGGACTACTTTTGCCTGTGAA 2928

RESULT 7  
AY078427 4006 bp mRNA linear ROD 01-JUL-2002  
LOCUS Mus musculus DNA cytosine methyltransferase 3b6 (Dnmt3b6) mRNA,  
DEFINITION complete cds; alternatively spliced.  
ACCESSION AY078427  
VERSION AY078427.1 GI:21655120  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Chen,T., Ueda,Y. and Li,E.  
TITLE Dnmt3a2 encoded by transcripts from an intronic promoter of Dnmt3a  
is expressed abundantly in ES cells and germ cells  
JOURNAL Unpublished  
AUTHORS Chen,T., Ueda,Y. and Li,E.  
TITLE Direct Submission  
JOURNAL Submitted (08-FEB-2002) Cardiovascular Research Center,  
Massachusetts General Hospital, 149 13th Street, Charlestown, MA  
02129, USA

FEATURES  
source  
1..4006  
/organism="Mus musculus"  
/strain="129S4/SvJae"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/map="84.0 cM"  
1..4006  
gene  
1..4006  
/gene="Dnmt3b6"  
269...4659  
/gene="Dnmt3b6"  
/functions="de novo methylation"  
/note="contains PHD and FWWP domains; alternatively  
spliced"  
/codon\_start=1  
/product="DNA cytosine methyltransferase 3b6"  
/protein\_id="AAL85481.1"  
/db\_xref="GI:21655121"  
/translation="MKGDSRLHNEEGASGYEECIIVNGNFSQSSDTKDAFPFVLE  
AICTPEVCTPETRRSSRLSREVSLNLTQDMTGDGRDDEVDVDDGSDILMPK  
LRTETKDTTRRSEPAVRTRHNGTSLSERQASPRITRGRQGRHHVQEPVPEFPATR  
SRRRRSSASTPWSPASVDPMEEVTPKSVSTPSVDLSQDQEGMDTTQVDAESRD  
GDSYEQDDKEFGIGDLVNGKI KGFSPMPAMVSWKATSKRQAMPGRMVQWFGDGKF  
SEI SADKLVALGFSOHFNKATFNKLVSRKAMYHTLEKARVAGKTFFSSPSGESLED  
KQPMLEWAGJGPKPTGIEGLKPNKKQPVNKSVRSDSRNLEPRRNRKSRRTTN  
DSASESPYPKRLTKNSYGGKDRGEDSERMASVETNNKNLEDRCLSCGSKNPFV  
PHLPEGLQCQCRDRLELFYMYDEDDGYQSVCTVCCGREGLLCSNTSCRCFCVEK  
LEVLVGATGAEDAKLQEPWPCYMCPLPQSRCHVLRKRDWNRQLDFFTTDPLEEFEP

PKLYPAIAKRRPIRVLSLFDGIATGYLVKELGIKVEKYIASBVCASIAVGTVKH  
EGQIKYVNDVRKIITKKNIENMGPPDLVIGSPCNDSNVNPKGLYSGTGRLPFFPY  
HLNLYTRPKGENDRPFWMFENVVMKVNKKDISRFACNPNVMDA1KVSAAHRARY  
FWNLFGNMRIIEGFFAHYTVDSNMGRCAROKLLGRSWSVPVIRHLFAPLKDYFACE"  
BASE COUNT 1001 a 1017 c 1075 g 913 t  
ORIGIN  
Alignment Scores: 8 56e-229 Length: 4006  
Pred. No.: 4223 50 Matches:  
Score: 92.67% Conservative: 0  
Percent Similarity: 92.67% Mismatches: 0  
Best Local Similarity: 92.67% Indels: 63  
Query Match: 10 Gaps: 1  
DB: 1  
US-09-720-086-6 (1-859) x AY078427 (1-4006)  
Qy 1 MetLysGlyAspSerArgHisLeuAenGluGluGlyAlaSerGlyTyrGluGluCys 20  
Db 269 ATGAAGGGAGACAGACACATCTGAATGAAGAAGAGGTGCCAGCGGTATGAGGAGTGC 328  
Qy 21 IleIleValAenGlyAsnPheSerAspGlnSerSerAspThrLysAspAlaProSerPro 40  
Db 329 ATTATCGTTAATGGAACTTTCAGTGACCACTCCTCAGACACGAAGGATGCTCCTCACCC 388  
Qy 41 ProValLeuGluAlaIleCysThrGluProValCysThrProGluThrArgGlyArg 60  
Db 389 CCAGTCTTGGAGGCAATCTGCACAGAGCCAGTCTGCACACAGAGACAGAGCCGACAG 448  
Qy 61 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuAenTyrThrGlnAspMet 80  
Db 449 TCAGCTCTCCGGCTGTCTAAGAGGAGGTCTCCAGCTTCTGAATACACGAGACATG 508  
Qy 81 ThrGlyAspGlyAspArgAspAspGluValAenGlyAsnGlySerAspIleLeuMet 100  
Db 509 ACAGAGATGAGACAGAGATGATGAAGTAGATGATGGAAATGGCTCTGATATTCCTAATG 568  
Qy 101 ProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaValArg 120  
Db 569 CCAAGCTCACCCGTGAGACCAAGGACACAGAGCCGCTCTGAAGACCCCGGTGTCCGA 628  
Qy 121 ThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArgAlaSerProArgIleThr 140  
Db 629 ACCCGACATAGCAATGGGACCTCCAGCTTGGAGAGGCAAGAGGCTCCCCCAGATCACC 688  
Qy 141 ArgGlyArgGlnGlyArgHisValGlnGluTyrProValGluPheProAlaThrArg 160  
Db 689 CGAGTCCGAGGCGCCGCCACCATGTGCAGGAGTACCTGTGGAGTTTCCGGCTACCAG 748  
Qy 161 SerArgArgArgAlaSerSerSerSerAlaSerThrProTrpSerSerProAlaSerVal 180  
Db 749 TCTCGAGACGTCGAGCATCTCTTCAGCAAGCAGCCATGGTCTATCCCTCGCAGCTC 808  
Qy 181 AspPheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGln 200  
Db 809 GACTTTCATGGGAAGAGTGACACCTAAGAGCGTCACTACCCCATCAGTTGACTTGAGCCAG 868  
Qy 201 AspGlyAspGlnGluGlyMetAspThrGlnValAenGluGluSerArgAspGlyAsp 220  
Db 869 GATGAGATCAGGAGGATGAGGATACCAACAGGTGGATGATGAGAGAGAGATGAGAGAC 928  
Qy 221 SerThrGluTyrGlnAspAspLysGluPheGlyIleGlyAspLeuValTrpGlyLysIle 240  
Db 929 AGCAGACAGTATCAGGATGATAAAGAGTTTGAAGTAGGTGACCTCGTGTGGGAAAGATC 988  
Qy 241 LysGlyPheSerTrpTrpProAlaMetValValSerTrpLysAlaThrSerLysArgGln 260  
Db 989 AAGGGCTTCTCCTGGTGGCTTGCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1048  
Qy 261 AlaMetProGlyMetArgTrpValGlnTrpPheGlyAspGlyLysPheSerGluIleSer 280  
Db 1049 GCCATGCCCGGAATGCCCTGGGTACAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1108

QY 281 AAlasplysleuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsn 300  
 Db 1109 GCTTAAACAAAGCTGGGCTCTGGGGCTGTTTCAGCCAGCACCTTAACTCTGGCTTCACTCAAT 1168  
 QY 301 LysleuValSerTyrArgGlyAlaMetCysThrIleuGluValAsnArgValArgAla 320  
 Db 1169 AAGCTGGTTCTTATAGAAAGCCATGTACACACTCTGGAGAAAGCCAGGGCTTCGACT 1228  
 QY 321 GlySerThrPheSerSerProGlyValSerLeuGluAspGlnLeuLysProMetLeu 340  
 Db 1229 GGCAAGACCTTCTCCAGAGACTCTGGAGATCATCTGAGAGACCAGCTGAAGCCATGCTG 1288  
 QY 341 GluTrpAlaHisGlyGlyPheLysProThrGlyIleGluGlyLeuValProAsnLysLys 360  
 Db 1289 GAGTGGCCCTCCGCTGGCTTCAAGCTCTGAGATCGAGGGCTTCAACCAACCAAGAG 1348  
 QY 361 GlnProValValAsnLysSerIleValArgArgSerAspSerArgAsnLeuGluProArg 380  
 Db 1349 CAAACAGTGTATATAGTCCAAAGGTGCTGCTCAGACATAGAGAACTTAGAACCCACAG 1408  
 QY 381 ArgArgGluAsnLysSerArgArgArgThrThrAsnAspSerAlaAlaSerGluSerPro 400  
 Db 1409 AGAGCGGAGAAACAAAGTCGAAAGACGACAAACCAATGACTCTGCTGCTTCTGAGTCCCC 1468  
 QY 401 ProProLysArgLeuLysThrAsnSerTyrGlyLysAspArgGlyGluAspGluGlu 420  
 Db 1469 CCACCCCAAGCCCTCAAGACAAATAGCTATGCGGGAGAGACCGAGGGAGATGAGAG 1528  
 QY 421 SerArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCys 440  
 Db 1529 AGCGGAGAAACGATGGCTCTGAAATCCACCAACAAAGGCAATCTGGAAGACCGCTGT 1588  
 QY 441 LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyGlyLeuLys 460  
 Db 1589 TTGCTCTGTGAAAGAAAGAACCCCTGTGCTTCCACCCCTCTTGAAGGTGGCTCTGT 1648  
 QY 461 GlnSerCysArgAspArgPheLeuGluLeuPheTyrMetTyrAspGluAspGlyTyrGln 480  
 Db 1649 CAGAGTGGCCGGAGTCGCTTCTAGAGCTTCTACATGATGATGAGAGCGGCTATCAG 1708  
 QY 481 SerTyrCysThrValCysCysGluGluValArgGluLeuLeuLysCysSerAsnThrSerCys 500  
 Db 1709 TCTCACTGACACGGTGTGTGTGAGGGCCGTGAACCTGCTGTCAGATMAACAAGCTGC 1768  
 QY 501 CysArgCysPheCysValGluCysLeuGluValLeuValGlyValAlaGlyThrAlaGluAsp 520  
 Db 1769 TGCAGATGCTTCTGTGTGAGATGTGTGAGGTGCTGTGGGCGGACAGCTGAGGAT 1828  
 QY 521 AlaLysLeuGluGluProTyrSerCysTyrMetCysLeuProGlnArgCysHisGlyVal 540  
 Db 1829 GCCAAGCTGACGAAACCTGTGAGCTGCTATGTGCTCCCTCAGCGCTGCATCGGGCTC 1888  
 QY 541 LeuArgArgArgGlyAspTyrAsnMetArgLeuGluAspPhePheThrThrAspProAsp 560  
 Db 1889 CTCGACGCGAGAAAGATTTGAAACATGCGCTGCAGAGCTTCTTCACTACATGATCTTAC 1948  
 QY 561 LeuGluGluPheGluProProLysLeuTyrProAlaIleProAlaAlaLysArgArgPro 580  
 Db 1949 CTGGAAGATTTGAGCCCAACAGTTGTACCAAGATTTCTGCGAGCCAAAGAGGGGCC 2008  
 QY 581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeu 600  
 Db 2009 ATTGAGAGCTCTGCTCTGTTTGTGATGGAATTCACACGGGTACTTGCTCTCAAGAGTTG 2068  
 QY 601 GlyIleLysValGluLysTyrIleAlaSerGluValCysAlaGluSerIleAlaValGly 620  
 Db 2069 GGATTTAAAGGAGAAAGATGATGCTCCCGAAGCTGTGAGAGTTCATGCTGTGGGA 2128  
 QY 621 ThrValLysHisGluGluGlnIleLysTyrValAsnAspValArgLysIleThrLysLys 640  
 Db 2129 ACTGTTAAAGCATGAAGGCCAATCATATATGTCAATGACGTCGGAAATATCAACAGAAA 2188  
 QY 641 AsnIleGluGluTyrGlyProPheAspLeuValIleGlyLysSerProCysAsnAspLeu 660

Db 2189 AATTATGAAGAGTGGGGCCCGTTCGACTTGTGATTTGTGTGAAGCCCATGAAATGATCTC 2248  
 QY 661 SerAsnValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhePheGlu 680  
 Db 2249 TCTAACGTCAATCTGTGCCCGCAAAAGTTTATATGAGGGCACAGGAAGGCTCTTTTTCGAG 2308  
 QY 681 PheTyrHisLeuLeuAsnTyrThrArgProLysGluGlyLysAspAsnArgProPhePheTyr 700  
 Db 2309 TTTTACCACTTGTGATTTATATCCCGCCCAAGAGGGCCACAACCGTCCATTTCTTCTGG 2368  
 QY 701 MetPheGluAsnValValAlaMetLysValAsnAspLysAspIleSerArgPheLeu 720  
 Db 2369 ATGTTCCAGAAATGTTGGCCATGAAAGTAAATGACAAAGAAAGCATCTCAAGATTCTCTG 2428  
 QY 721 AlaCysAsnProValMetIleAspAlaIleLysValSerAlaHisArgAlaArgTyr 740  
 Db 2429 GCATGTAAACCAAGTGAATGATGATGCCATCAAGGTGTCTCTGCTCACAGGGCCGGGTAC 2488  
 QY 741 PheTrpGlyAsnLeuProGlyMetAsnArgProValMetAlaSerLysAsnAspLysLeu 760  
 Db 2489 TTTCTGGGTAACTTACCCCGGAATGAAC----- 2515  
 QY 761 GluLeuGluAspCysLeuGluPheSerArgThrAlaLysLeuLysValGlnThrIle 780  
 Db 2515 ----- 2515  
 QY 781 ThrThrLysSerAsnSerIleArgGlnGlyLysAsnGlnLeuPheProValValMetAsn 800  
 Db 2515 ----- 2515  
 QY 801 GlyLysAspAspValLeuTyrCysThrGluLeuGluArgIlePheGlyPheProAlaHis 820  
 Db 2516 -----AGGATCTTGGCTTCCCTGCTCAC 2519  
 QY 821 TyrThrAspValSerAsnMetGlyArgGlyAlaArgGlnLysLeuGluArgSerTyr 840  
 Db 2540 TACACGAGACGTGTCCAACATGCGCGCGCGCGCTCAGAAAGTGTGAGGAGGTCTCTGG 2599  
 QY 841 SerValProValIleArgHisLeuPheAlaProLeuLysAspTyrPheAlaCysGlu 859  
 Db 2600 AGGTACCGGTCAATCAGACACTGTTTGGCCCTTGAAGACATACTTGGCTGTGAA 2656  
 RESULT 8  
 AF151975  
 LOCUS AF151975 4034 bp mRNA linear ROD 08-JUN-2000  
 DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 7  
 (Dnmt3b) mRNA, complete cds.  
 ACCESSION AF151975  
 VERSION AF151975.1 GI:8347134  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 4034)  
 Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and  
 Shen, Y.  
 Cloning of full-length Dnmt3b cDNA and its alternative splicing  
 isoforms in mouse embryonic tissue  
 JOURNAL  
 REFERENCE  
 2 (bases 1 to 4034)  
 Unpublished  
 Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and  
 Shen, Y.  
 Direct Submision  
 Submitted (17-MAY-1999) Department of Biochemistry and Molecular  
 Biology, Institution of Basic Medical Sciences, Chinese Academy of  
 Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,  
 P. R. China  
 FEATURES  
 source  
 location/Qualifiers  
 1..4034  
 /organism="Mus musculus"  
 /strain="KM"

```
/db_xref="taxon:10090"
/dev_stage="8-9 day old embryo"
1..4034
/gene="Dnmt3b"
CDS
297..2687
/gene="Dnmt3b"
/function="methylates cytosine in DNA"
/note="alternatively spliced"
/codon_start=1
/product="DNA cytosine-specific methyltransferase isoform 7"
```

```
/protein_id="AAF74521.1"
/db_xref="GI:8347135"
/translators="MKGDSRHLNEBEGASGVVECIIVNGNFSQSDTKDAPSPVLE
AICTEPCVTPETGRSSRLSKREVSLNLTQDMDGDRDDEVDGNGSDILMPK
LRTKETDTRRSSEPAVTRHNSGTSSLEBRQASPRITRGRQGHVQVYPPVEFPATR
SRRRASSTASPTWSSPASVDFMEETPKSVSTPSVLSQDQOQEGMDTTOYDAESIY
GDSTEYODDFEFGIGLVWVKIKGFSPWPAWVVMVKATSKROAMPGRWVQWFGDGKF
SEIADKLVALGLFQSHFNALFNKLVSVRKAMVHTLEKARVRAGTKFSSSGESLED
OLKPMLEWAGGKFKPTIEGLKPNKKQPVNKSVRSDSRNLEPRRRNCKSKKNPVS
DSASESPPPKRLKNTSYGDKRGDESESRERMASEVTNNKNLEDRCLSCRCFCVEE
PHPLPEGGLQACDRFLFVYMYDEGQVSYCTVCCGRELLELLCSNTTCDPDLSEFEP
LEVLAGCTADEAKLOEPWSCYMLPQRCGVLRRRKDMNMLQDFFTDPLEFEP
PKLPAIPAARRFIRVLSLFDGIATGLVLKELGIVKVEYIASEVCASSIAGTVKH
EGQIKYNDVRKLTKNIREWGPFDLVIGGSPCNLDSNVNPAKGLYEGTGRLEFEY
HLLAYTRPKEDNRPFPWFENVMVKYNDKDLISRELACNPVMDAILKVSABHARY
FWGNLPQNMRIFFGPAHYTDVSNMGRGAKQLLGRSNSVPIRHLFAPLKYDFACE"
```

BASE COUNT 991 a 1042 c 1097 g 904 t

ORIGIN

## Alignment Scores:

```
Pred. No.: 7.82e-228 Length: 4034
Score: 4206.50 Matches: 794
Percent Similarity: 92.43% Conservative: 0
Best Local Similarity: 92.43% Mismatches: 2
Query Match: 91.64% Indels: 63
DB: 10 Gaps: 1
```

US-09-720-086-6 (1-859) x AF151975 (1-4034)

```
Qy 1 MetLysGlyAspSerArgHisLeuAsnGluGluGluGluValAsnSerGlyTyrGluGluCys 20
Db 297 ATGAAGGAGACACAGACATCTGAATGAAGAAGAGGGTGCACCGGGTATGAGGAGTGC 356
Qy 21 IleIleValAsnGlyAsnPheSerAspGlnSerSerAspThrIleAspAlaProSerPro 40
Db 357 ATTATCGTTAATGGGNACTTCAGTGACCAGTCTCCACACAGAGAGTGTCTCCACCC 416
Qy 41 ProValLeuGluAlaIleCysThrCluProValCysThrProGluThrArgGlyArgArg 60
Db 417 CCAAGTCTTGAGGCAATCTGCACAGAGCCAGTCTGCACACACAGACACAGAGCGCGAGG 476
Qy 61 SerSerSerArgLeuSerIleArgGluValSerSerIleuLeuAsnTyrThrGlnAspMet 80
Db 477 TCAAGCTCCCGGCTGTCTAAGAGGAGGTCTCCAGCGCTTCTGAATTCACACAGAGACATG 536
Qy 81 ThrGlyAspGlyAspArgAspAluValAspAspGlyAsnGlySerAspIleuMet 100
Db 537 ACAGGAGATGGACAGAGATGATGAATGATGATGGAAATGGCTCTGATATTTCTAATG 596
Qy 101 ProLysLeuThrArgGluThrIleAspThrArgThrArgSerGluSerProAlaValArg 120
Db 597 CCAAAGCTCACCGGTGAGACCAAGACACAGGACGCGTCTGAAAGCCCGGCTGTCCGA 656
Qy 121 ThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArgAlaSerProArgIleThr 140
Db 657 ACCCGACATAGCAATGGGACCTCCAGCTTTGAGAGGCAAGAGCGCTCCCCCAGAAATCAC 716
Qy 141 ArgGlyArgGluGlyArgHisIleValGluCluTyrProValGluPheProAlaThrArg 160
Db 717 CGAGGTCCGACAGGGCGCCACCATGTCGACGAGTACCTGTGGAGTTTCCGGCTTACCAGG 776
Qy 161 SerArgArgArgAlaSerSerSerAlaSerThrProTrpSerSerProAlaSerVal 180
```

```
Db 777 TCTCGAGACGCTCGAGCATCGTCTTTCAGCAAGCAGCGCCATCGTCTATCCCTGCCACGCTC 836
Qy 181 AspPheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGln 200
Db 837 GACTTCATGGAGAGGTGACACACTAAGAGCGTCAGTACCCCATCAGTTGACTTGAGCCAG 896
Qy 201 AspGlyAspGlnGluGlyMetAspThrThrGlnValAspAlaGluSerArgAspGlyAsp 220
Db 897 GATGAGATCAGGAGGTATGGATACACACAGGTGGATCGAGAGACATATATGAGAC 956
Qy 221 SerThrGluTyrGlnAspAspLysGluPheGlyIleGlyAspLeuValTrpGlyLysIle 240
Db 957 AGCAGAGATATCAGGATGATAAAGAGTTTGAAGTAGTACCTCGTGTGGGAAAGATC 1016
Qy 241 LysGlyPheSerTyrTrpProAlaMetValValSerTyrLysAlaThrSerLysArgGln 260
Db 1017 AAGGGCTTCTCTCGTGGCTGCCATCGTGGTGTCTCTGGAAAGCCACCTCCAAAGCCACAG 1076
Qy 261 AlaMetProGlyMetArgTyrValGlnTyrPheGlyAspGlyLysPheSerGluIleSer 280
Db 1077 GCCATGCCCGCAATGCCCTGGGTACAGTGGTGTGGTATGCAAGTCTTCTGAGATCTCT 1136
Qy 281 AlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsn 300
Db 1137 GCTGACAAACTGGTGGCTCTGGGGCTGTTCACCCAGCACCTTAACTCTGGCTACCTTCAAT 1196
Qy 301 LysLeuValSerTyrArgLysAlaMetTyrHisThrLeuGluLysAlaArgValArgAla 320
Db 1197 AAGCTGGTCTTCTTATAGAGGCCATGTACACACTCTGGAGAAAGCCAGGTTCCAGCT 1256
Qy 321 GlyLysThrPheSerSerProGlyGluSerLeuGluAspGlnLeuLysProMetLeu 340
Db 1257 GGCAGAGACCTTCTCCAGCAGTCTCTGAGAGTCACTGGAGGACACAGCTGAAAGCCATGCTG 1316
Qy 341 GluTrpAlaHisGlyGlyPheLysProThrGlyIleGluGlyLeuLysProAsnLysLys 360
Db 1317 GAGTGGGCCACCGTGGCTTCAAGCCTACTGGGATCGAGGCGCTCAAAACCAACAAGAA 1376
Qy 361 GlnProValAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProArg 380
Db 1377 CAACAGTGGTAAATAGTCCAAAGTGCCTGCTTCAGACAGTAGGAACTTAGAACCCAGG 1436
Qy 381 ArgArgGluAsnLysSerArgArgThrAsnAspSerAlaLaserGluSerPro 400
Db 1437 AGACGGGAGAACAAAAGTCCGAGACGCGCACCAACCAATGACTCTGCTGCTTCTGAGTCCCC 1496
Qy 401 ProProLysArgLeuLysThrAsnSerTyrGlyLysAspArgGlyGluAspGluGlu 420
Db 1497 CCACCCAAAGCGCTTCAAGACCAATAGCTATGGGGAAGGACCGAGGGAGGATGAGGAG 1556
Qy 421 SerArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCys 440
Db 1557 AGCCGAGAACCGATGGCTTCTGAAGTCACCAACAAGGCAATCTGGAAAGACCGCTGT 1616
Qy 441 LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyGlyLeuCys 460
Db 1617 TTGTCTCTGTGAAAGAAAGAACCTCTGTGCTTCCACCCCTCTTTGAGGGTGGGCTCTGT 1676
Qy 461 GlnSerCysArgAspArgPheLeuGluLeuPheTyrMetTyrAspGluAspGlyTyrGln 480
Db 1677 CAGAGTTCGCCGGATCGCTTCTTAGAGCTCTTCTACATGTATGATGAGGACGCTATCAG 1736
Qy 481 SerTyrCysThrValCysCysGluArgGluLeuLeuLeuCysSerAsnThrSerCys 500
Db 1737 TCCTACTGCACCGTGTCTGTGAGGCGCTGAACTGCTGCTGTGCTGCTGCTGCTGCTGCTG 1796
Qy 501 CysArgCysPheCysValGluCysLeuGluValLeuValGlyAlaGlyThrAlaGluAsp 520
Db 1797 TGCAGATGCTTCTGTGTGGAGTGTCTGGAGGTGTCTGTGTGGCGGACGACAGCTGAGGAT 1856
Qy 521 AlaLysLeuGlnGluProTrpSerCysTyrMetCysLeuProGlnArgCysHisGlyVal 540
```

Db 1857 GCCAAGCTGCAGAACCTGAGAGCTGCTATATGTGCTCCTCCTCAGCGCTGCATGGGCTC 1916  
 Qy 541 LeuArgArgArgLyAspTyrAspMetArgLeuGlnAspPhePheThrThrAspProAsp 560  
 Db 1917 CTCGACGACGACGAGAAAGATTGGAAACATGGCGCTCGCAAGACTTCTTCACTAGATCTGCAC 1976  
 Qy 561 LeuGlnGlnPheGlnPheProProLyLeuTyrProAlaIleProAlaAlaLyArgArgPro 580  
 Db 1977 CTGGAAGAGATTGGACCCACCCAGATTGTACCCACCAATTCCTGAGCCGCAAGAGAGGCC 2036  
 Qy 581 IleArgValLeuSerSerLeuPheAspGlyIleAlaThrGlyTyrLeuValIleuLeuGlnLeu 600  
 Db 2037 ATTGAGCTCCCTGCTCTCTGTTGATGAGATTCCACAGGGGTACTTGCTGCTCAGAGAGTTG 2096  
 Qy 601 GlyIleuValGlnuValGlyTyrIleAlaSerGlnuValCysAlaGlnuSerIleAlaValGly 620  
 Db 2097 GGTATTAAAGTGAAGAAAGTACATTGCTCCGAGAGCTGTGACAGATGCTCATGCTGTGGGA 2156  
 Qy 621 ThrValLySerGlnGlnGlnIleuLyTyrValAsnAspValArgLySileThrLyLeuLy 640  
 Db 2157 ACTGTTAAGCATGGAAGGCCACATCAATATGTCAATGACGTCGCCGAAATCACAAGAGAA 2216  
 Qy 641 AsnIleGlnGlnuTyrPGLYProPheAspLeuValIleGlyGlySerProCysAsnAspLeu 660  
 Db 2217 AATATTGAAGAGTGGGGCCCTGCTGACTGTGATGTGGAGGCCGCAATGATGATCTC 2276  
 Qy 661 SerAsnValAsnProAlaArgLyGlyLeuTyrGlnGlyThrGlyArgLeuPhePheGln 680  
 Db 2277 TCTAAGCTGCATCTGCTGCCGCAAGGTTTATATAGGGCACAGAGAGCTCTTCTTCAG 2336  
 Qy 681 PheTyrHisLeuLeuAsnTyrThrArgProLyGlnGlyAspAsnArgProPhePheTyr 700  
 Db 2337 TTTTACCACTTGCATGATTATATACCCGCCCAAGAGGGCGCAACCTGCTCTTCTCG 2396  
 Qy 701 MetPheGlnuAsnValIleAlaMetLyValAsnAspLyLeuAspIleSerPhePheLeu 720  
 Db 2397 ATGTTGCGAAGTGTGTGGCCATGAAAGTGAATGACAAAGAACATCTCAAGATCTCTG 2456  
 Qy 721 AlaCysAsnProValMetIleAspAlaIleLyValSerAlaAlaHisArgAlaArgTyr 740  
 Db 2457 GCATGTAAACCCAGATGATGATGATGCCATCAAGGCTGTGCTGTCTCACAGAGCCCGGTAC 2516  
 Qy 741 PheTyrGlnuLeuProGlyMetAsnArgProValMetAlaSerLyAsnAspLyLeu 760  
 Db 2517 TTTCTGGGTAACTTACCCGGAATGAC----- 2543  
 Qy 761 GlnuLeuGlnuAspCysLeuGlnuPheSerArgThrAlaLyLeuLyValGlnThrIle 780  
 Db 2543 ----- 2543  
 Qy 781 ThrThrLySerAsnSerIleArgGlnGlyLyAsnGlnLeuPheProValIleMetAsn 800  
 Db 2543 ----- 2543  
 Qy 801 GlyLyAspAspValLeuTyrCysThrGlnuLeuGlnuArgIlePheGlyPheProAlaHis 820  
 Db 2544 ----- 2544  
 Qy 821 TyrThrAspValSerAsnMetGlyArgGlyAlaArgGlnuLyLeuLeuGlyArgSerTyr 840  
 Db 2568 TACACGGAGCTGTCCAAATGATGGCGCGCGCGCTCAGAAAGCTGCTGGAGAGGCTCTGG 2627  
 Qy 841 SerValProValIleArgHisLeuPheAlaProLeuLyAspTyrPheAlaCysGln 859  
 Db 2628 AGTGTACCGTCACTACACACCTGTTTCCCTTGAAGAGACTTGTGCTGTGA 2684  
 RESULT 9  
 LOCUS AF151972  
 DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 4  
 (Dnmt3b) mRNA, complete cds.  
 ACCESSION AF151972  
 VERSION AF151972.1 GI:8347125

**KEYWORDS**  
 SOURCE Mus musculus  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
**REFERENCE**  
 AUTHORS Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and Shen, Y.  
 TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing isoforms in mouse embryonic tissue  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 4149)  
 AUTHORS Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and Shen, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular Biology, Institution of Basic Medical Sciences, Chinese Academy of Medical Sciences, Dong Dan San Tiao 5, Beijing city 100005, P.R.China  
**FEATURES**  
 source  
 1..4149  
 /organism="Mus musculus"  
 /strain="KM"  
 /db\_xref="taxon:10090"  
 /dev\_stage="8-9 day old embryo"  
 1..4149  
 /gene="Dnmt3b"  
 412..2802  
 /gene="Dnmt3b"  
 /function="methylates cytosine in DNA"  
 /note="alternatively spliced"  
 /codon\_start=1  
 /product="DNA cytosine-specific methyltransferase isoform 4"  
 /protein\_id="AA074518.1"  
 /db\_xref="GI:8347126"  
 /translation="MRGDSRHLLNEEGAGYEECTIVNGNFSQSSDTKADAPPVLE  
 AICTEPVCTPETERGRSSRLSKREVSLNLYADMTGDGRDDEVDGSDILMPK  
 LTRBTDTRTRESRPVTRHSNGTSLRORASPRITRGRORHNVQEPVFPAR  
 SRRRRSSASSTPMSPSVADPMEVTPSPSVPSVDSLDGQEGMDTQVAAESLY  
 GDSDEVODDKREGIDLVNCKIKGSPWAMPVYSKATSKROMPCRMWOMRGDGKF  
 SETSADKLVALGLFSQHFNDATFNKLVSTRAMTHTLEKRVAGKTFSSPESLED  
 QLSAPLEMAHGGFKPTGIEGLKPKNKQPVNNKSVKRSOSRNLPRRREKRSRTTN  
 DSAAPSPPKPKLTNSYGGKDEBESREPMASVTNNKNGLEBRCLSCGKKNPV  
 LFLPFGGGLCOSCRDRFLFLFYMDGVOYSTVCCEGELLGNTSCRCFVEEC  
 LEVLVAGTAEDAKLOEPMSCYMLPDRCHGVLRRRKDMRLDQFFTDPDLEFEP  
 PKLYPAIPAKRRPRIVLSFDGIATGVLVXELGIVKERTIASVCAESIANCTVH  
 EGQIKTVNDVRKTIKKNIEBGFPLVIGSPONDLSVNPARKGLREGTGRLPFEFY  
 HLNTYTRPREGDNRPFPMFENVAKVNDKDIISPLACNPVMDAIVKSAHARAY  
 FMGNLPGMNRIFGFPAYHTDVSNMGRGAKQLLGRSWSVPIVHLPAFLDYFACE"  
**BASE COUNT** 1013 a 1082 c 1123 g 931 t  
**ORIGIN**  
**Alignment Scores:**  
 Pred. No.: 8.09e-228 Length: 4149  
 Score: 4206.50 Matches: 794  
 Percent Similarity: 92.43% Conservative: 0  
 Best Local Similarity: 92.43% Mismatches: 2  
 Query Match: 91.64% Indels: 63  
 DB: 10 Gaps: 1  
 US-09-720-086-6 (1-859) x AF151972 (1-4149)  
 Qy 1 MetLyGlnuLyAspSerArgHisLeuAsnGlnuGlnuGlnuLyAlaSerGlyTyrGlnuGlnu 20  
 Db 412 ATGAAGGAGACACACACACTGTGAATGAAGAGAGGTCACAGGGGTATAGAGAGAGCG 471  
 Qy 21 IleIleValaengLyAsnAspSerArgHisLeuAsnGlnuGlnuGlnuLyAlaSerGlyTyrGlnuGlnu 40  
 Db 472 ATTATCGTTAATGGAACTTCAAGACCAAGTCTCACAACAGAAAGATGCTCCTCACCC 531  
 Qy 41 ProValLeuGlnuAlaIleCysThrGlnuProValCysThrProGlnuThrArgGlyArgArg 60



Db 532 CCMGCTTGGAGGCAATCTGCACAGACCAGTCTGCACACCAGACACGAGCGCCAGG 591  
Qy 61 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTyrThrGlnAspMet 80  
Db 592 TCNAGCTCCCGGCTGTCTAAGAGGGAGGTCTCCAGCTTCTGAAATTACACGAGGACATG 651  
Qy 81 ThrGlyAspGlyAspArgAspGluValAspAspGlyAsnGlySerAspIleLeuMet 100  
Db 652 ACAGGAGATGGAGACAGAGATGATGAAGTAGATGATGGGAATGGCTCTGATATTCTAATG 711  
Qy 101 ProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaValArg 120  
Db 712 CCAAAGCTCACCGGTGAGACCAAGGACACGAGACGCGCTCTGAAGCCCGGCTGTGCCGA 771  
Qy 121 ThrArgHisSerAsnGlyThrSerSerLeuGluThrGlnArgAlaSerProArgIleThr 140  
Db 772 ACCCGACATAGCAATGGGACCTCCAGCTTGGAGAGCAAGAGAGCTCCGCCAGAAATCAC 831  
Qy 141 ArgGlyArgGlnGlyArgHisValGlnGluTyrProValGluPheProAlaThrArg 160  
Db 832 CGAGGTTCGACAGGGCGCCACCATGTGCAGAGTACCTCTGTGGAGTTTCCGGCTTACCAGG 891  
Qy 161 SerArgArgArgAlaSerSerSerAlaSerThrProTrpSerSerProAlaSerVal 180  
Db 892 TCTCGAGACGTGAGAGCATCTCTCTTCAGCAGACACGCCATGTCATCCCTGCCAGGTC 951  
Qy 181 AspPheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGln 200  
Db 952 GACTTTCATGGAAAGAGTGACACCTAAGAGCGTCAGTACCCTCATCAGTTGACTTGGCCAG 1011  
Qy 201 AspGlyAspGlnGluGlyMetAspThrThrGlnValAspAlaGluSerArgAspGlyAsp 220  
Db 1012 GATGGAGATCAGAGGGGTATGGATACCACACAGGTGGATGCAGAGAGCATATATGGAGAC 1071  
Qy 221 SerThrGluTyrGlnAspAspLysGluPheGlyIleGlyAspLeuValTrpGlyLysIle 240  
Db 1072 AGCACAGAGTATCAGAGATATAAGAGTTTGGAAATAGGTGACCTCTGTGGGGGAAAGATC 1131  
Qy 241 LysGlyPheSerTrpTrpProAlaMetValValSerTrpLysAlaThrSerLysArgGln 260  
Db 1132 AAGGGCTTCTCTGTGGTGGCTGCCATGGTGTGTCTCTGGAAAGGCCACCTCCCAAGCGACAG 1191  
Qy 261 AlaMetProGlyMetArgTrpValGlnTrpPheGlyAspGlyLysPheSerGluIleSer 280  
Db 1192 GCCATGCCCGGAATGGCTGGGTACAGTGGTTTGGTATGGCAAGTTTCTGAGATCTCT 1251  
Qy 281 AlaAspLysLeuValAlaLeuGlyLeuPheSerSerGlnHisPheAsnLeuAlaThrPheAsn 300  
Db 1252 GCTGACAACTGGTGGCTCTGGGGCTGTTCAGCCAGCAGCTTTAATCTGGCTACCTTCAAT 1311  
Qy 301 LysLeuValSerTyrArgLysAlaMetTyrHisThrLeuGluLysAlaArgValArgAla 320  
Db 1312 AAGCTGGTTTCTTATAGGAGGCCATGTACCACACTCTGGAGAAAGCCAGGTTCCAGCT 1371  
Qy 321 GlyLysThrPheSerSerProGlyLysSerLeuAspGluAspGlnLeuLysProMetLeu 340  
Db 1372 GGCAAGACCTTCTCCAGCAGTCTGGAGAGTCACTGGAGGACGAGTGAAGCCCATGCTG 1431  
Qy 341 GlnTrpAlaHisGlyGlyPheLysProThrGlyIleGluLysLeuPheProAsnLysLys 360  
Db 1432 GAGTGGGCCCCACGGTGGCTTCAAGCTTACTGGGATCGAGGGCTCAAAACCCCAAGAAAG 1491  
Qy 361 GlnProValValAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProArg 380  
Db 1492 CAACCACTGGTTAATTAAGTCGAAGGTGCTGTTTCAGACAGTAGTAGGAACCTTAGAACCCAGG 1551  
Qy 381 ArgArgGluAsnLysSerArgArgThrThrAsnAspSerAlaAlaSerGluSerPro 400  
Db 1552 AGACCGAGAAACAAAGTTCGAAGAGCGCACCAACCAATGACTCTGCTGCTTCTGAGTCCCC 1611  
Qy 401 ProProLysArgLeuLysThrAsnSerTyrGlyGlyAspArgGlyLysGluGlu 420  
Db 1612 CCACCCAGGCGCTCAAGACAAATAGTATGGCGGAGGACCTTGGCGGAGGAGTGGAGG 1671

Qy 421 SerArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCys 440  
Db 1672 AGCCGAGAACCGATGGCTTCTGAAGTCAACCAACAAAGGCAATCTGGAAGACCGCTGT 1731  
Qy 441 LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyLysLeuCys 460  
Db 1732 TTGTCTCTGTGGAAGAAGAACCCCTGTGTCTTCCACCCCTCTTTGAGGGTGGGCTCTGT 1791  
Qy 461 GlnSerCysArgAspArgPheLeuGluLeuPheTyrMetTyrAspLysAspGlyTyrGln 480  
Db 1792 CAGAGTTTCCCGGATCGCTTCTTAGAGCTCTTCTACATGATGATGAGGACGCGCTATCAG 1851  
Qy 481 SerTyrCysThrValCysGluGluClyArgGluLeuLeuLeuCysSerAsnThrSerCys 500  
Db 1852 TCCCTACTGACCGGTGTGTGAGGCGGTGACTGCTGTGCGAGTAAACAAGCTGC 1911  
Qy 501 CysArgCysPheCysValGluCysLeuGluValLeuValGlyAlaGlyThrAlaGluAsp 520  
Db 1912 TGCAGATGCTTCTGTGTGAGGTGCTGTGAGGTGCTGTGGCGCAGGACAGCTGAGGAT 1971  
Qy 521 AlaLysLeuGlnGluProTrpSerCysTyrMetCysLeuProGlnArgCysHisGlyVal 540  
Db 1972 GCCAAGCTGCAGGAACCCCTGGAGCTGTATATGTGCTCTCCCTCAGCGCTGCCATGGGTC 2031  
Qy 541 LeuArgArgArgLysAspTrpAsnMetArgLeuGlnAspPheThrThrAspProAsp 560  
Db 2032 CTCCGACGCGAGGAAGATTGGAAACATGCGCTGCAAGACTTCTTCACTACTGATCCTGAC 2091  
Qy 561 LeuGluGluPheGluProProLysLeuTyrProAlaIleProAlaLysArgArgPro 580  
Db 2092 CTGGAGAAATTTGAGCCACCCCAAGTTGTACCCAGCAATTCCTCAGCCAAAGAGGCGCC 2151  
Qy 581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeu 600  
Db 2152 ATTAGAGTCTCTCTCTGTGTTGATGAATTCGACCGGGTACTTGTGTCTCAAGAGTTG 2211  
Qy 601 GlyIleLysValGluLysTyrIleAlaSerGluValCysAlaGluSerIleAlaValGly 620  
Db 2212 GGTATTAAAGTGAAGATACATTCCTCCCAAGTCTGTGCAGAGTCCATCGCTGTGGGA 2271  
Qy 621 ThrValLysHisGluGlyGlnIleLysTyrValAsnAspValArgLysIleThrLysLys 640  
Db 2272 ACTGTTAAGCATGAAGCCAGATCAAAATATGTCAATGACGTCCGGAATAATCAAGAAA 2331  
Qy 641 AsnIleGluGluTrpGlyProPheAspLeuValIleGlyGlySerProCysAsnAspLeu 660  
Db 2332 AATATTGAAGAGTGGGGCCCGTTCCGACTTGGTGTATGGTGAAGCCCATGCAATGATCTC 2391  
Qy 661 SerAsnValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPheGlu 680  
Db 2392 TCTAAGCTCAATCTGCTCCGCAAGGTTTATATGAGGCGACAGGAAGGCTTCTTCTCGAG 2451  
Qy 681 PheTyrHisLeuLeuAsnTyrThrArgProLysGluGlyAspAsnArgProPhePheTrp 700  
Db 2452 TTTTACCCTTGTGGAATTTATACCCGCCCAAGGAGGCGGACACCGTCCATTTCTTGG 2511  
Qy 701 MetPheGluAsnValValAlaMetLysValAsnAspLysLysAspIleSerArgPheLeu 720  
Db 2512 ATGTTTCGAGAATGTGTGGCCATGAAGTGAATGACAAGAAGACATCTCAAGTTCTCTG 2571  
Qy 721 AlaCysAsnProValMetIleAspAlaIleLysValSerAlaHisArgAlaArgTyr 740  
Db 2572 GCATGTAAACCGAGTATGATCGATGCCATCAAGGTGTCTGTCTGCTCAGAGGCGCGGTAC 2631  
Qy 741 PheTrpGlyAsnLeuProGlyMetAsnArgProValMetAlaSerLysAsnAspLysLeu 760  
Db 2632 TTCTGGGTAACTACCCGGAATGAAC----- 2658  
Qy 761 GluLeuGlnAspCysLeuGluPheSerArgThrAlaLysLeuLysValGlnThrIle 780  
Db 2658 ----- 2658



QY 781 ThrThrIysSerAseSerIleArgGlnGlyLysAnGlnLeuPheProValIValMetAsn 800  
 Db 2658 ----- 2658  
 QY 801 GlyLysAspAseValLeuTrpCysThrGluLeuGlnArgIlePheGlyPheProAlaHis 820  
 Db 2659 ----- 2662  
 QY 821 TyrThrAspValIserAsnMetGlyArgGlyAlaArgGlnLysLeuLeuGlyArgSerTrp 840  
 Db 2663 TACACGACGCTGTCCACATATGCGCGCGCGCGCGCGCTGCAAGAGCTGTGGCAGAGCTCCGG 2742  
 QY 841 SerValProValIleArgHisLeuPheAlaProLeuLysAspTyrPheAlaCysGln 859  
 Db 2743 AGGTACCGGTCATCAGACACCTGTTGCCCTTGAAAGGACTACTTGTCTGTGAA 2799  
 REF 10 AF068628 3946 bp mRNA linear ROD 06-DEC-1999  
 AF 28 AF068628  
 LOCUS DEFINITION Mus musculus DNA cytosine-5 methyltransferase 3B3 (Dnmt3b) mRNA,  
 alternatively spliced, complete cds.  
 ACCESSION AF068628 GI:6449473  
 VERSION AF068628.2 GI:6449473  
 KEYWORDS Mus musculus.  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Bakayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 3946)  
 AUTHORS Okano, M., Xie, S., and Li, E.  
 TITLE Cloning and characterization of a family of novel mammalian DNA  
 (cytosine-5) methyltransferases  
 JOURNAL Nat. Genet. 19 (3), 219-220 (1998)  
 MEDLINE 98324766  
 PUBMED 9662389  
 REFERENCE 2 (bases 1 to 3946)  
 AUTHORS Xie, S., Okano, M., and Li, E.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,  
 Charlestown, MA 02129, USA  
 REFERENCE 3 (bases 1 to 3946)  
 AUTHORS Okano, M., Chijiwa, T., Sasaki, H., and Li, E.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,  
 Charlestown, MA 02129, USA  
 COMMENT Sequence update by submitter  
 On Nov 18, 1999 this sequence version replaced gi:3527983.  
 FEATURES  
 SOURCE  
 1. 3946  
 /location/Qualifiers  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="2"  
 /map="84.0 cM"  
 1. 3946  
 /gene="Dnmt3b"  
 1. 3946  
 /note="similar to EST sequences deposited in GenBank  
 Accession Numbers AA116694, AA119979, AA177277, AA210568,  
 AA407106, and AA575617"  
 269. 2599  
 /gene="Dnmt3b"  
 /function="de novo DNA methylation"  
 /note="alternatively spliced product; contains Cys-rich  
 region"  
 /codon\_start=1  
 /product="DNA cytosine-5 methyltransferase 3B3"  
 /protein\_id="AAC40180.2"  
 /db\_xref="GI:6449474"  
 /translation="MKGDSRHINEEGASGVEECITVNGNPSDSSPTKXAPSPVLE  
 AICPEYCTPETGRKRSSSRKRSKREVSLLNTYTDMDGDDRDDEVDNGSDILNPK  
 LTRERKTRRSSEPAVTRHSNGTSLERQASPRITRGRQGRHHVQEVPEVPATR  
 SRRRASSASPTWSPASVDFMEVTPKSVSPVLSQDGDGDMDTTQVDAESRD  
 misc\_feature  
 CDS

BASE COUNT 983 a 1006 c 1056 g 901 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 3,7e-222 Length: 3946  
 Score: 4105.50 Matches: 776  
 Percent Similarity: 90.34% Conservative: 0  
 Best Local Similarity: 90.34% Mismatches: 83  
 Query Match: 89.44% Indels: 83  
 DB: 10 Gaps: 2  
 US-09-720-086-6 (1-859) x AF068628 (1-3946)  
 QY 1 MethylsGlyAspSerArgHisLeuArgGlnGluGlnAlaSerGlyTyrGluGlnCys 20  
 Db 269 ATGAGAGGAGACAGACACATCTGAATGAAGAGAGGCTCCAGCGGTATGAGAGATGC 328  
 QY 21 IleIleValAnGlyAsnPheSerAspGlnSerSerAspThrLysAspAlaProSerPro 40  
 Db 329 ATTATCGTAAATGGAACTTCAAGACACGCTCTCAACACGAAAGATGCTCCCTCAACC 388  
 QY 41 ProValLeuGlnAlaIleCysThrGluProValCysThrProGluThrArgGlyArgArg 60  
 Db 389 CCAAGCTTGAAGGCAATCTCCACAGACGCTCCGACACACAGACACAGAGCCGACAG 448  
 QY 61 SerSerSerArgLeuSerLysArgGlnValSerSerLeuLeuAsnTyrThrGlnAspMet 80  
 Db 449 TCAAGCTCCCGGCTGTCTTAAGAGGAGGCTCCAGCTTGAATTACAGCGACAGATG 508  
 QY 81 ThrGlyAspGlyAspArgAspAspGlnValAspAspGlyAnGlySerAspIleLeuMet 100  
 Db 509 ACAGAGATGAGACACAGAGATGATGAATGATGATGATGATGATGATGATGATGATG 568  
 QY 101 ProLysLeuThrArgGlnTyrTrpLysAspThrArgThrArgSerGlnSerProAlaValArg 120  
 Db 569 CCAAGCTCAACCTGAGACCAAGACACAGACGCGCTCTGAAGACCGCGCTGTCCGA 628  
 QY 121 ThrArgHisSerAnGlyThrSerSerLeuGlnArgGlnArgAlaSerProArgIleThr 140  
 Db 629 ACCGACATAGCAATGAGACCTCCAGCTTGAGAGCGAAAGAGCTCCCGACAGATCACC 688  
 QY 141 ArgGlyArgGlnGlyArgHisValGlnGlnGlnTyrTrpProValGluPheProAlaThrArg 160  
 Db 689 CGAGGTGGCAGAGGCGCCACCATGTCCAGAGTACCTTGAGATTCCGGCTACACAGG 748  
 QY 161 SerArgArgArgArgAlaSerSerSerSerAlaSerThrProTrpSerSerProAlaSerVal 180  
 Db 749 TCTGGAGACGTCGAGCATGTCTTTCAGCAAGACAGCGCATGTATCTCCCTGCAAGCTC 808  
 QY 181 AspPheMetGluGlnValIleTrpProLysSerValSerThrProSerValAspLeuSerGln 200  
 Db 809 GACTTCATGGAAGAGAGACCTTAAGAGCGTCACTCCCATCAGTTGACTTGAAGCAG 868  
 QY 201 AspGlyAspGlnGlnGluMetAspThrThrGlnValAspAlaGluSerArgAspGlyAsp 220  
 Db 869 GATGAGATCAGAGGAGTATGATACCAACAGGTGATCCAGAGACAGACAGATGAGAC 928  
 QY 221 SerThrGluTyrGlnAspAspLysGluPheGlyIleGlyAspLeuValITPGLYLye 240  
 Db 929 AGCACAAGATATCGATGATAAAGATTTTGAAATAGTACCTCGTGTGGGAAAATC 988  
 QY 241 LysGlyPheSerTrpTrpProAlaMetValSerTrpLysAlaThrSerLysArgGln 260

Db 989 AAGGGCTTCTCTGGTGGCTGCCATGGTGGTGTCTGTGAAAGCCACCTCCCAAGCGACAG 1048  
Qy 261 AlaMetProGlyMetArgTyrValGluThrPheGlyAspGlyValPheSerGluIleSer 280  
Db 1049 GCCATGCCCGGAATCGCTGGGTACAGTGGTTGGTGATGGCAAGCTTTCTGAGATCTCT 1108  
Qy 281 AlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsn 300  
Db 1109 GCTGACAAACTGGTGGCTCTGGGGCTGTTCAGGCAGCACTTTAAATCTGGCTACCTTCAAT 1168  
Qy 301 LysLeuValSerTyrArgLysAlaMetTyrHisThrIleuGluLysAlaArgValArgAla 320  
Db 1169 AAGCTGGTTTCTTTATAGGAAGCCCATGTACCACTCTGGAGAAAGCCAGGGTTCGAGCT 1228  
Qy 321 GlyLysThrPheSerSerProGlyGlySerLeuGluAspGlnLeuLysProMetIleu 340  
Db 1229 GGCAAGACCTTCTCCAGCAGCTCTGGAGAGTCACTGGAGGACAGCTGAAGCCCATCTCTG 1288  
Qy 341 GluThrAlaHisGlyGlyPheLysProThrGlyIleGluGlyLeuLysProAsnLysLys 360  
Db 1289 GAGTGGGCCCCACGGTGGCTTCAAGCCTACTGGGATCGAGGGCTCAAAACCCAAACAAGAG 1348  
Qy 361 GlnProValValAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProArg 380  
Db 1349 CAACCA----- 1354  
Qy 381 ArgArgGluAsnLysSerArgArgArgThrThrAsnAspSerAlaAlaSerGluSerPro 400  
Db 1355 -----GAGAACAAAGTTCGAGAGCGCACCAACNATGACTCTGCTGCTTCTGAGTCCCCC 1408  
Qy 401 ProProLysArgLeuLysThrAsnSerTyrGlyLysAspArgGlyGluAspGluGlu 420  
Db 1409 CCACCACAGCGCTCAAGACAAATAGCTATGGCGGAGAGCCGAGGGGAGATGAGGAG 1468  
Qy 421 SerArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCys 440  
Db 1469 AGCCGAGAACCGATGGCTTCTGAAGTCACCAACAAGGGCAATCTCGAAGACCGCTGT 1528  
Qy 441 LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyGlyLysCys 460  
Db 1529 TTGCTCTGTGGAAGAAGAACCCCTGTGTCTCTCCACCCCTCTTTGAGGGGTGGGCTCTGT 1588  
Qy 461 GlnSerCysArgAspArgPheLeuGluLysPheTyrMetTyrAspGluAspGlyTyrGln 480  
Db 1589 CAGAGTTGCCGGGATCGCTTCTAGAGCTCTTCTACATGTATGATGAGGAGCGGTATCAG 1648  
Qy 481 SerTyrCysThrValCysCysGluGlyValArgGluLeuLeuLysCysSerAsnThrSerCys 500  
Db 1649 TCCTACTGCACCGTGTGTGTGAGGGCGGTGAACCTGCTGTCAGTAAACACAAGCTGC 1708  
Qy 501 CysArgCysPheCysValGluCysLeuGluValLeuValGlyValAlaGlyThrAlaGluAsp 520  
Db 1709 TGCAGATGCTTCTGTGTGAGTGTCTGGAGTCTGTGGTGGCGCAGGCACAGCTGAGGAT 1768  
Qy 521 AlaLysLeuGlnGluProThrSerCysTyrMetCysLeuProGlnArgCysHisGlyVal 540  
Db 1769 GCCAAGCTGCAGAACCCCTGAGCTGTATGTGCTCTCCCTCAGCGCTGCATGGGGTC 1828  
Qy 541 LeuArgArgLysAspTyrAsnMetArgLeuGluAspPhePheThrThrAspProAsp 560  
Db 1829 CTCGACGCGAGAAAGATTGGAACATCGCTCGCAAGACTTCTCAGTACTGATCTCTGAC 1888  
Qy 561 LeuGluGluPheGluProProLysLeuTyrProAlaIleProAlaLysLysArgArgPro 580  
Db 1889 CTGGAAGAATTGAGCCACCCCAAGTTGTACCCAGCAATTCCTGCGAGCCAAAGGAGGCC 1948  
Qy 581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLysGluLeu 600  
Db 1949 ATTAGAGTCTCTGTCTGTGTTGATGGAATTGCAACGGGGTACTTGTGTCTCAAGGAGTTG 2008  
Qy 601 GlyIleLysValGluLysTyrIleAlaSerGluValCysAlaGluSerIleAlaValGly 620  
Db 2009 GGTATTAAAGTGGAAAAGTACATTGCCTCCGAAAGTCTGTCCAGAGTCCATCGTGTGGGA 2068

Qy 621 ThrValLysHisGluGlyGlnIleLysTyrValAsnAspValArgLysIleThrLysLys 640  
Db 2069 ACTGTTAAGCATGAAGCCAGATCAAAATATGTCATGACGTCCGGAATATCACCAGAAA 2128  
Qy 641 AsnIleGluGluTyrGlyProPheAspLeuValIleGlyGlySerProCysAsnAspLeu 660  
Db 2129 AATATTGAAGAGTGGGCCCTTCGACTTGGTGATTTGGTGAAGCCCATGCAATGATCTC 2188  
Qy 661 SerAsnValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPheGlu 680  
Db 2189 TCTAACGTCATCTGTCGCCCAAGGTTTATATGAGGACACAGAAAGGCTCTTCTTCGAG 2248  
Qy 681 PheTyrHisLeuLeuLeuAsnTyrThrArgProLysGluGlyAspAsnArgProPhePheTyr 700  
Db 2249 TTTTACCACCTTGTCTGAATATATACCCGCCCAAGGAGGCGCACACCGTCCATCTTCTGG 2308  
Qy 701 MetPheGluAsnValAlaMetLysValAsnAspLysLysAspLysSerArgPheLeu 720  
Db 2309 ATGTTCCGAGAAATGTTGGCCATGAAAGTGAATGACAAAGAACATCTCAAGATTCTCTG 2368  
Qy 721 AlaCysAsnProValMetIleAspAlaIleLysValSerAlaAlaHisArgAlaArgTyr 740  
Db 2369 GCATGTAAACCCAGTGTATGATCGATCCATCAAGGTGTCTGCTCACAGGGCCCGGTAC 2428  
Qy 741 PheTyrGlyAsnLeuProGlyMetAsnArgProValMetAlaSerLysAsnAspLysLeu 760  
Db 2429 TTTCTGGGTTAACTTACCCGGAATGAAC----- 2455  
Qy 761 GluLeuGlnAspCysLeuGluPheSerArgThrAlaLysLeuLysValGlnThrIle 780  
Db 2455 ----- 2455  
Qy 781 ThrThrLysSerAsnSerIleArgGlnGlyLysAsnGlnLeuPheProValValMetAsn 800  
Db 2455 ----- 2455  
Qy 801 GlyLysAspAspValLeuTyrCysThrGluLeuGluArgIlePheGlyPheProAlaHis 820  
Db 2456 -----AGGATCTTCGGCTTCCTCGCTCAC 2479  
Qy 821 TyrThrAspValSerAsnMetGlyArgGlyAlaArgGlnLysLeuGlyArgSerTyr 840  
Db 2480 TACACGACGTGTCCAACATGGCGCGCGCCCTCAGAAGCTGTGGGAGCTCTGG 2539  
Qy 841 SerValProValIleArgHisLeuPheAlaProLeuLysAspTyrPheAlaCysGlu 859  
Db 2540 AGTGTACCGGTATCAGACACCTGTTTGGCCCTTGAAGGACTACTTTGCTGTGAA 2596

RESULT 11  
AF151976 3974 bp mRNA linear ROD 08-JUN-2000  
LOCUS  
DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 8  
(Dnmt3b mRNA, complete cds).  
ACCESSION AF151976  
VERSION AF151976.1 GI:8347136  
KEYWORDS  
SOURCE Mus musculus.  
ORGANISM Mus musculus  
REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. (bases 1 to 3974)  
AUTHORS Yin, B., Chen, Y.T., Zhu, M., Luo, Y.J., Zhu, N., Xu, S.C., Wu, G.Y. and Shen, Y.  
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing isoforms in mouse embryonic tissue  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3974)  
AUTHORS Yin, B., Chen, Y.T., Zhu, M., Luo, Y.J., Zhu, N., Xu, S.C., Wu, G.Y. and Shen, Y.  
TITLE Direct Submission  
JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular Biology, Institution of Basic Medical Sciences, Chinese Academy of



```
QY 501 CysArgCysPheCysValGluCysLeuGluValLeuValGluValAlaGlyThrAlaGluAsp 520
Db 1737 TGCAGATGCTTCTGTGTGAGTGCTGTGAGGTGCTGTGCGCCAGGCACAGCTGAGGAT 1796
QY 521 AlalaValGluGlnGluProTrpSerCysTyrMetCysLeuProGlnAlaCysHisGlyVal 540
Db 1797 GCCAGCTGCAGGACCCCTGAGCTGTATATGTGCTCTCCCTCAGCGGTGCCATGGGGTC 1856
QY 541 LeuArgArgArgLysAspTrpAsnMetArgLeuGlnAspPhePheThrAspProAsp 560
Db 1857 CTCGCGAGGAGAAAGATTGGAACATGCGCTGCAAGACTTCTTACACTGATGATCCCTGAC 1916
QY 561 LeuGluGluPheGluProProLysLeuTyrProAlaIleProAlaAlaLysArgArgPro 580
Db 1917 CTGGAAGAATTTGAGCCACCCCAAGTTGTACCCAGCAATTCCTGAGCCAAAGAGAGCCC 1976
QY 581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeu 600
Db 1977 ATTAGAGTCTGCTCTGTTGTTGATGAATTGCAACGGGGTACTTGTGTCTCAAGAGATTG 2036
QY 601 GlyIleLysValGluLysTyrIleAlaSerGluValCysAlaGluSerIleAlaValGly 620
Db 2037 GGTATTAAAGTGAAGAGTACATTCCTCCGAAGTCTGTGACAGGTCCATCGCTGTGGGA 2096
QY 621 ThrValLysHisGluGlnIleLysTyrValAsnAspValArgLysIleThrLysLys 640
Db 2097 ACTGTTAAGCATGAAGCCAGATCAATATGTCATGACGTCCGGAATATCCCAAGAA 2156
QY 641 AsnIleGluGluTrpGlyProPheAspLeuValIleGlySerProCysAsnAspLeu 660
Db 2157 AATATTGAAGAGTGGGGCCGTTGACTGTGATTTGTTGAGGAGCCCATGATGATCTC 2216
QY 661 SerAsnValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhePheGlu 680
Db 2217 TCTAAGCTCAATCTCCCGCCAAAGTTTATATAGGGGACACAGGAAGCTCTTCTTCGAG 2276
QY 681 PheTyrHisLeuLeuAsnTyrThrArgProLysGluGlyValAspAsnArgProPheTyr 700
Db 2277 TTTTACCACCTGCTGAATATACCCGCCCAAGAGGGGCAACCGCTCCATCTTCTCGG 2336
QY 701 MetPheGluAsnValAlaMetLysValAsnAspLysLysAspIleSerArgPheLeu 720
Db 2337 ATGTTCCAGAAATGTTGTGGCCATGAAGTGAATGACAGAAAGACATCTCAAGATCTCTG 2396
QY 721 AlaCysAsnProValMetIleAspAlaIleLysValSerAlaAlaHisArgAlaArgTyr 740
Db 2397 GCATGTAACCCAGTCATGATCCATCAAGGTGCTGCTGCTCACAGGGCCCGGTAC 2456
QY 741 PheTrpGlyAsnLeuProGlyMetAsnArgProValMetAlaSerLysAsnAspLysLeu 760
Db 2457 TTTCTGGGTAACTACCCGGAATGAC----- 2483
QY 761 GluLeuGlnAspCysLeuGluPheSerArgThrAlaLysLeuLysValGlnThrIle 780
Db 2483 ----- 2483
QY 781 ThrThrLysSerAsnSerIleArgGlnGlyLysAsnGlnLeuPheProValValMetAsn 800
Db 2483 ----- 2483
QY 801 GlyLysAspAspValLeuTrpCysThrGluLeuGluArgIlePheGlyPheProAlaHis 820
Db 2484 -----AGGATCTTCGGCTTCCCTGCTCAC 2507
QY 821 TyrThrAspValSerAsnMetGlyArgGluValArgGlnLysLeuGlyArgSerTrp 840
Db 2508 TACACGGAGCTGTCAACATGGCGCGCGCGCGCTCAGAACTGCTGGCAGGCTCTGG 2567
QY 841 SerValProValIleArgHisLeuPheAlaProLeuLysAspTyrPheAlaCysGlu 859
Db 2568 AGTGTAACCGGTATCAGACACCTGTTTGGCCCCCTTGAAGGACTACTTTGCTCTGTGAA 2624
RESULT 12
```

```
AF1511971
LOCUS      4089 bp      mRNA      linear      ROD 08-JUN-2000
DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 3
ACCESSION AF1511971
VERSION   AF1511971
KEYWORDS  AF1511971.1 GI:8347122
SOURCE    Mus musculus.
ORGANISM  Mus musculus.
REFERENCE 1 (bases 1 to 4089)
AUTHORS   Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
           Shen,Y.
TITLE      Cloning of full-length Dnmt3b cDNA and its alternative splicing
           isoforms in mouse embryonic tissue
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 4089)
AUTHORS   Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
           Shen,Y.
TITLE      Direct Submission
JOURNAL   Submitted (17-MAY-1999) Department of Biochemistry and Molecular
           Biology, Institution of Basic Medical Sciences, Chinese Academy of
           Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,
           P.R.China
FEATURES   Location/Qualifiers
            source          1..4089
                                /organism="Mus musculus"
                                /strain="KM"
                                /db_xref="taxon:10090"
                                /dev_stage="8-9 day old embryo"
            gene            1..4089
            CDS             412..2742
                                /gene="Dnmt3b"
                                /genes="Dnmt3b"
                                /function="methylates cytosine in DNA"
                                /notes="alternatively spliced"
                                /codon_start=1
                                /product="DNA cytosine-specific methyltransferase isoform
                                3"
                                /protein_id="AAF74517.1"
                                /db_xref="GI:8347123"
                                /translation="MKGDSRLNKEEGAGYECIIVNGNFSQSSDTKDAPSPVLE
                                AICTEPTCTPTGRSSRLSKREVSLLNTYQDMTGDGDDDDVDDGNSDILMP
                                LTRTKDTRTSSEPAVTRHSNGTSLERQASPRIITRGROGRHHVDEYVPEVDAES
                                SRRRASSASTPMSVPASVDPMFEVTPKSVTPSVDSQDQSGMDTTOVDAES
                                GDSTEYQDDKEFGIDLVGKIKGFSWPMVYVSKATSKROAMFGMBWVCGDGKF
                                SETSADKLVALGLSQHENLATFNKLYSVKMYHTLEKARVAGKTPESSPGSLSD
                                QLKPMLEWAGGFKPTGLEGLKPKKOPENKSRPTTNDASAESPPEKRLKTNLYGG
                                KRGDEDESERMASSEVTNKGLEKRCISCKKAPYSFPLFEGCLQCSDRLDEL
                                FMYDDEGYSCTYCCGRELILCSNTSCRCFCVECLEVLVAGTADKLGSPMS
                                CYNLPQRCHGVLRKDKNNRGLQDFITDPLESEFEPKLTPLPAARRRPIRVLSL
                                FDGIATGILVLKELGKIKVETIASSEVCAESIAVGVKHEGQIKYVNDVRKITKKNIE
                                WGPFDLIVIGSPENDLSNVNPARKGLYEGTGRLLFFEFYHLNLTTPKSGDNRPFWMF
                                ENVVANKVNDKDISRFIACNPFVMDIAIKVSAARARYFWMGLPGMNRIFGFPAAHYTD
                                VSNMGRGARQKLGRSMSPVIRHLFAPLKDYPACE"
BASE COUNT 995 a 1071 c 1104 g 919 t
ORIGIN
Alignment Scores:
Pred. No.:      3..5e-221      Length:      4089
Score:          4088.50      Matches:    774
Percent Similarity: 90.10%      Conservative: 0
Best Local Similarity: 90.10%      Mismatches: 2
Query Match:    89.07%      Indels:      83
DB:             10      Gaps:        2
US-09-720-086-6 (1-859) x AF1511971 (1-4089)
QY 1 MetLysGlyAspSerArgHisLeuAsnGluGluGlyAlaSerGlyTyrGluGlyCys 20
Db 412 ATGAGAGGAGACACACACATCTGATGAAGAGAGGCGTCCACCGGGTATGAGGATGC 471
```

QY 21 IleIleValaAngIyaAmPheSerAaspGlnSerSerAaspThrIlyAspAlaProSerPro 40  
 Db 472 ATTATCGTTAATGGAAATCTTCAAGACCAAGCTCTCCAGACAGAAAGAAAGCTCCCTCAACC 531  
 QY 41 ProValLeuGluAlaIleCysThrGluProValCysThrProGluThrArgGlyArgArg 60  
 Db 532 CCAAGCTTTGGAGGCAATTCGACAGAGCCAGTCTGACACACCAAGAGCCAGAGCCGACAG 591  
 QY 61 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTyThrGlnAspMet 80  
 Db 592 TCAAGCTCCCGGCTGTCTAAGAGGAGGCTCCAGCCTTCTGAATTACCAAGACGACATG 651  
 QY 81 ThrGluAspGlyAspArgAspArgGluValAspAspGlyAsnGlySerAspIleLeuMet 100  
 Db 652 ACAGAGATGAGACAGAGATGATGAATGATGGAAATGGCTCTGATATTTCTAATG 711  
 QY 101 ProValLeuThrArgGluThrLysAspThrArgThrArgSerArgLysProAlaValArg 120  
 Db 712 CCAAGCTCACCCGTTGAGACCAAGAGACACAGAGCCGCTCTGAAAGCCCGGCTGTCGA 771  
 QY 121 ThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArgAlaSerProArgIleThr 140  
 Db 772 ACCGACATACATGCAATGGGACCTCCAGCTTGGAGAGGCAAAAGCTCCCCAGAAATCAC 831  
 QY 141 ArgGluArgGlnGlyArgHisIleValGlnGluTyProValGluPheProAlaThrArg 160  
 Db 832 CGAGGTGGAGGAGGCGCCGACCATGTGCAGAGATACCTGTGAGATTTCCGGCTACAG 891  
 QY 161 SerArgArgArgArgAlaSerSerSerAlaSerThrProTyrPseSerProAlaSerVal 180  
 Db 892 TCTGGAGACCTGACAGTCGTCTTCAAGCAAGACGCGCATGTCACTCCCGCCAGCCGTC 951  
 QY 181 AspPheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGln 200  
 Db 952 GACTTCATGAGAAAGATGACACCTAAGCGCTCAGTACCCTCATGATGACTTGAGCCAG 1011  
 QY 201 AspGluAspGlnGluGlyMetAspThrThrGlnValAspAlaGluSerArgAspGlyAsp 220  
 Db 1012 GATGAGATTCAGAGGAGTATGATACCAACAGGTGATGACAGAGCATATATGAGAAC 1071  
 QY 221 SerThrGluTyArgGlnAspAspLysGluPheGlyIleGlyAspLeuValTyrGlyIle 240  
 Db 1072 AGCAGAGATTCAGAGATGAATAAAGATTGGAATGATGACCTCGTGGGAAAGATC 1131  
 QY 241 LysGluPheSerTyrTyrProAlaMetValIleSerTyrLysAlaThrSerIysArgGln 260  
 Db 1132 AAGGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1191  
 QY 261 AlaMetProGlyMetArgTyrValGlnTyrPheGlyAspGlyLysPheSerGluIleSer 280  
 Db 1192 GCCATGCCCGGAATGCGCTGGGTACAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1251  
 QY 281 AlaAspLysLeuValAlaLeuGlyLeuPheSerGlnIlePheAsnLeuAlaThrPheAsn 300  
 Db 1252 GCTGACAAACGTGGCTCTGGGGCTGTTCAAGCCAGCACTTAACTGGCTACCTTCAAT 1311  
 QY 301 LysLeuValSerTyrArgLysAlaMetTyrHisThrLeuGluLysValAlaArgValAlaArg 320  
 Db 1312 AAGCTGTTCTTATAGGAAGGCCATGTAACCACTCTGGAGAAAGCCAGGGTGTGAGCT 1371  
 QY 321 GlnTyThrPheSerSerSerProGlyLysLeuSerLeuGluAspGlnLeuLysProMetLeu 340  
 Db 1372 GGCAAGACCTTCTCCAGAGCTCTGAGAGTCACTGAGAGACCACTGAAGCCATGCTG 1431  
 QY 341 GlnTyrAlaHisGlyGlyPheLysProThrGlyIleGluGlyLeuLysProAsnLys 360  
 Db 1432 GAGTGGGCGCCAGGTGGCTTCAAGCCTCTGGGATCGAGGGCTCAAAACCCCAAGAG 1491  
 QY 361 GlnProValValAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProArg 380  
 Db 1492 CAACCA----- 1497  
 QY 381 ArgArgGluAsnLysSerArgArgThrThrAsnAspSerAlaIleSerGluSerPro 400

Db 1498 -----GAGAACAAAGTCCGAGAGCGACCAACATATCTGCTGCTTCTGATCTCCCC 1551  
 QY 401 ProProLysArgLeuLysThrAsnSerTyrGlyLysAspArgGlyGluAspGluGlu 420  
 Db 1552 CCACCCAAAGCCCTCAAGACAAATAGCTATGCGGGAAGAGCCAGGAGGAGATGAGAG 1611  
 QY 421 SerArgLysArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCys 440  
 Db 1612 AGCCGAGAACCGATGGCTCTGAAGTCAACCAACAAAGGCAATCTGGAAGACCGGCTGT 1671  
 QY 441 LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyLysCys 460  
 Db 1672 TTGCTCTGGAAAGAAAGAACCTGTGTCTTCAACCCCTCTTGGAGGGGCTGCTGT 1731  
 QY 461 GlnSerCysArgAspArgPheLeuGluLeuPheTyrMetTyrAspGluAspGlyTyrGln 480  
 Db 1732 CAGAGTCCCGGATCCGCTCTCTAGACTCTTCACTATGATGATGAGAGCGCTATCAG 1791  
 QY 481 SerTyrCysThrValCysCysGluGlyArgGluLeuLeuLeuCysSerAsnThrSerCys 500  
 Db 1792 TCTTACTGCAACGTGTCTGTGAGAGGCGGTGAATGCTGTGTGTCAGATACCAAGCTGC 1851  
 QY 501 CysArgCysPheCysValGluCysLeuGluValLeuValGlyAlaGlyThrAlaGluAsp 520  
 Db 1852 TGCAGATGCTTCTGTGTGAGTGTCTGAGGTGCTGTGGGCGGACAGCGCTGAGGAT 1911  
 QY 521 AlaLysLeuGlnGluProTyrPseCysTyrMetCysLeuProGlnArgCysHisGlyVal 540  
 Db 1912 GCCAAGCTGACAGAACCTTGAGCTGCTATATGTGCTCTCCACAGCGTGCATAGGGGTC 1971  
 QY 541 LeuArgArgArgLysAspTyrAsnMetArgLeuGlnAspPheThrThrAspProAsp 560  
 Db 1972 CTCGACGACAGAAAGATTGGAACATGGCGCTGCAAGCTTCTTCACTAGATCTTGAC 2031  
 QY 561 LeuGluGluPheGluProProLysLeuTyrProAlaIleProAlaAlaLysArgArgPro 580  
 Db 2032 CTGGAAGATTGAGCCACCCAGCTTGACCCAGCAATCTCTGACGCCAAAGAGAGGCC 2091  
 QY 581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeu 600  
 Db 2092 ATTAGATCTGTCTCTGTGTATGATGAATGCACAGGGGATCTTGAGTCTCAAGAGATTG 2151  
 QY 601 GlyIleLysValGluLysTyrIleAlaSerGluValCysAlaGluSerIleAlaValGly 620  
 Db 2152 GGTATTAAAGTGAAGAAAGTATGCTCCGAGTCTGTGACAGTGCATCGCTGTGGA 2211  
 QY 621 ThrValLysHisGluGlyGlnIleLysTyrValAsnAspValArgLysIleThrLysLys 640  
 Db 2212 ACTGTTAAGCATGAAGGCCAGATCAAAATATGTCAATGACGTCGGAATAATCACCAAGAA 2271  
 QY 641 AsnIleGluGluTyrProGlyProPheAspLeuValIleGlyGlySerProCysAsnAspLeu 660  
 Db 2272 AATATTGAAGGTGGGGCCCTGTGACTGTGTGATGGTGGAAACCCCATGGAAATGATCTC 2331  
 QY 661 SerAsnValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhePheGlu 680  
 Db 2332 TCTTAACGTCAATCTGCCCGGAAGGTTTATATAGGGCACAAGAGCTCTTCTTGAG 2391  
 QY 681 PheTyrHisLeuLeuAsnTyThrArgProLysGluGlyAspAsnArgProPhePheTyr 700  
 Db 2392 TTTTAAACACTTGTAATTATTAACCGCCCAAGAGGGCGCAACCGCTCATTTCTTCTGG 2451  
 QY 701 MetPheGluAsnValValAlaMetLysValAsnAspLysLysAspIleSerArgPheLeu 720  
 Db 2452 ATGTTGAGATGTTGGCCCATGAAGTGAATGACAAAGAAACATCTCAAGATTCTGTG 2511  
 QY 721 AlaCysAsnProValMetIleAspAlaIleLysValSerAlaAlaHisArgAlaArgTyr 740  
 Db 2512 GCATGTAAACCAAGTATGATGATGCCATCAAGGTGTCTGTCTCACAGGGCCCGGTAC 2571  
 QY 741 PheTyrGlyAsnLeuProGlyMetAsnArgProValMetAlaSerLysAsnAspLysLeu 760

Db 2572 TTCTGGGTAACCTACCCGAATGAAC----- 2598  
Qy 761 GluLeuGlnAspCysLeuGluPheSerArgThrAlaLysLeuLysValGlnThrIle 780  
Db 2598 ----- 2598  
Qy 781 ThrThrLysSerAsnSerIleArgGlnGlyAsnGlnLeuPheProValValMetAsn 800  
Db 2598 ----- 2598  
Qy 801 GlyLysAspValLeuTrpCysThrGluLeuGluArgIlePheGlyPheProAlaHis 820  
Db 2599 -----AGGATCTTCGGCTTCCCTGCTCAC 2622  
Qy 821 TyrThrAspValSerAsnMetGlyArgGlyAlaArgGlnLysLeuGlyArgSerTrp 840  
Db 2623 TACACGAGCGTGTCAACATCGCGCGCGCGCGCGTCAAGCTCTCGGCGAGTCTCTGG 2682  
Qy 841 SerValProValIleArgHisLeuPheAlaProLeuLysAspTyrPheAlaCysGlu 859  
Db 2683 AGTGTACCGTGCATCAGACACCTGTTTGCCTTGAAGGACTACTTTGCCTGTGAA 2739

## RESULT 13

AF156488 4145 bp mRNA linear PRI 05-SEP-1999  
LOCUS Homo sapiens DNA cytosine-5 methyltransferase 3 beta 1 (DNMT3B)

DEFINITION mRNA, complete cds.

ACCESSION AF156488

VERSION AF156488.1 GI:5823167

KEYWORDS

SOURCE Homo sapiens.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 4145)

Xie S., Wang Z., Okano M., Nogami M., Li Y., He W.W., Okumura K.

and Li E.

Cloning, expression and chromosome locations of the human DNMT3

gene family

Gene 236 (1), 87-95 (1999)

10433969

2 (bases 1 to 4145)

Xie S. and Li E.

Direct Submission

Submitted (03-JUN-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,

Charlestown, MA 02129, USA

Location/Qualifiers

1..4145

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="20"

/map="between D20S182 and D20S106; 39.9-50.2 cM"

1..4145

/gene="DNMT3B"

115..2676

/gene="DNMT3B"

/note="de novo DNA methyltransferase"

/codon\_start=1

/product="DNA cytosine-5 methyltransferase 3 beta 1"

/protein\_id="AAB53063.1"

/db\_xref="GI:5823168"

/translations="MKGDTRHLNGEEDAGGREGSILVNGACSDSSDPPILAIRTP

VRNRNNSVSRERHRSRPRSTRGRQGRNHVDESPVFPKLFRETRRSSEPA

DKSPGLDILVMGKIKGFSMWPMVVMKATKROAMGMRWQWFGDCKFSEVADK

LVALGLPSQHFNLATKGLVSRKAMVHAKERVRAGTTPSSPGDSLDDOLKPMLE

WAHGGVPGTIEGLKPNNTQPVVNVKVRKAGSKLESREKVRKRRVADDSATDV

CPAPKRLKTCYNNKADGDEDOQREOMADYANNKSSLEDGLSCGRKNKPVSPFLF

EGGQCICRDRFELFPTMDDDGSTICTVCCGREGILLNCNFCRCRCFCVCELEVLV

GTGTAAAKLQEPWSCITMCLPQRCHGVLRKRWNRWLRQAFSTDSUTGLSEAPKLPPA

IPAAARRPIRLVLPGLATGLVLKELGIVKGVYASEVCEESIAVGTVRHEGNIKI

VNDVRNITTKNTIEEMGPFDDLVIGGSPCNDLSNVNPARKGLYEGLTGRLLRPFYHLNYS  
RPKEGDRPFFPMFNVAMVGVDRDI SRFLCNPVMIDAI KVSAAHRARYFPMGNLP  
GMNRPVIAASKNDKLEQCLVEYNRIAKLKKVQITITTKSNSIKQGNQLPVPVMMGSKED  
VLMTLELRIFGFPVHYTDVSNMGRGARQKLLGRSNVSPVIRHLFAPLKDYFACE"

BASE COUNT 1041 a 1083 c 1096 g 925 t

ORIGIN

Alignment Scores:

Pred. No.: 1..26e-202 Length: 4145

Score: 3759.00 Matches: 717

Percent Similarity: 88.63% Conservative: 55

Best Local Similarity: 82.32% Mismatches: 69

Query Match: 81.90% Indels: 30

DB: 9 Gaps: 10

US-09-720-086-6 (1-859) x AF156488 (1-4145)

Qy 1 MetLysGlyAspSerArgHisLeuAsnGluGluGluGlyAlaSerGlyTyrGluGluCys 20

Db 115 ATGAAGGGAGACACACCGCATCTCAATGGAGAGGAGGACGCCGGGAGGAAGACTCG 174

Qy 21 IleIleValAsnGlyAsnPheSerAspGlnSerSerAspThrLysAspAlaProSerPro 40

Db 175 ATCTCTGTCACAGGGGCTCGACGGACCATCTCTCGAC-----TCGCC 219

Qy 41 ProValLeuGluAlaIleCysThrGluProValCysThrProGluThrArgGlyArg 60

Db 220 CCAATCTTGGAGGCTATC-----CCACCCCGGAGATCAGAGCCCGAAGA 264

Qy 61 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTyrThrGlnAspMet 80

Db 265 TCAAGCTCGACACTCTCAAGAGGGAGGTGTCACACTCTGCTAAAGTACACACAGGACTTG 324

Qy 81 ThrGlyAspGlyAspArgAspAspGluValAspAspGlyAsnGlySerAspIle---Leu 99

Db 325 ACAGCGCATGCGCAGCGG-----GAAGATGGGATGCTCTGACACCCCGAGT 372

Qy 100 MetProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaVal 119

Db 373 ATGCCAAAGCTCTTCCCGGAA-----ACCAGGACTCTTTCAGAAAGCCCGAGTGT 423

Qy 120 ArgThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArgAlaSerProArgIle 139

Db 424 CGAACTCGAAATAACAACAGTGTCTCCAGCCGGAGAGGACACAGGCTTCCCGCTTCC 483

Qy 140 ThrArgGlyArgGlnGlyArgHisValGlnGlyTyrProValGluPheProAlaThr 159

Db 484 ACCCGAGCCCGGAGGCGCCCAACCATGTGGACAGTCCCGCTGGAGTTCCCGGTACC 543

Qy 160 ArgSerArgArgArgAlaSerSerSerAlaSerThrProTyrSerSerProAlaSer 179

Db 544 AGGTCCCTGAGACCGCGGCGCAACAGCATCGGAGGAAAGCCATGCGCTCTCCCTCCAGC 603

Qy 180 -----ValAspPheMetGluGluVal-----ThrProLysSer 190

Db 604 TCTTACCTTACCATCGACCTTCACAGACGACACAGAGGACACATGGGAGCCCGCAGAGC 663

Qy 191 ValSerThrProSerValAspLeuSerGlnAspGlyAspGlnGluGlyMetAspThrThr 210

Db 664 AGCAGTACCCCTACGCCCGCTAGCCAGGACAGCCAGGAGGCGCATGGAGTCCCGC 723

Qy 211 GlnValAspAlaGluSerArgAspGlyAspSerThrGluTyrGlnAspAspLysGluPhe 230

Db 724 CAGGTGGAGGACAGACAGTGGAGATGAGACAGTTCAGAGTATCAGATGGAGAGGAGTTT 783

Qy 231 GlyIleGlyAspLeuValTrpGlyLysIleLysGlyPheSerTrpTrpProAlaMetVal 250

Db 784 GGATAGGGGACCTCTGTGTGGGAAAGATCAAGGCTTCTCTGTGTGGCCCGCATGGTG 843

Qy 251 ValSerTrpLysAlaThrSerLysArgGlnAlaMetProGlyMetArgTrpValGlnTrp 270

Db 844 GTGTCTTGGAAAGGCCCTCCAAAGCAGAGGCTATGTCTGGCATGCGGTGGGTCCAGTGG 903

QY 271 PheGlyAspGlyLysPheSerGluIleSerAlaAspLysLeuValAlaLeuGlyLeuPhe 290  
 Db 904 TTGGGCGATGGCAAGTCTTCGAGGTCTCGACACAACAGGGGCACTGGGGCTGTC 963  
 QY 291 SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyr 310  
 Db 964 AGCCAGACACTTAATTGGCCACTTCAATAGCTCTCTCTTATCGAAGAGCCATGTAC 1023  
 QY 311 HisThrLeuGluLysAlaArgValArgAlaGlyLysThrPheSerSerProGlyGlu 330  
 Db 1024 CATGCTCTGGAGAAAGCTAGGGGTGCGAGCTGCAAGACCTTCCCGAGAGCCCTGGAGAC 1083  
 QY 331 SerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGlyPheLysProThr 350  
 Db 1084 TCATTGGAGGACCGAGCTGAAACCATGTGGAGTGGGCCCGGGGGCTTCAGACCCCT 1143  
 QY 351 GlyIleGluGlyLeuLysProAsnLysGlnProValAlaAsnLysSerLysValArg 370  
 Db 1144 GGGATCGAGGGCTCCAAACCAACAACACGCAACCGAGTTAATAGTCAGAGGTGCGT 1203  
 QY 371 ArgSerAspSerArgAsnLeuGluProArgArgLysLeuAsnLysSerArgArgThr 390  
 Db 1204 CGTCAGAGCAGTAGAAGATTAGATCAAGAATAACGAGAACAGACTCGAAGACGACA 1263  
 QY 391 ThrAsnAspSerAlaAlaSerGlu--SerProProLysArgLysThrAsnSer 409  
 Db 1264 GCTGACCACTCAGCCACTCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1323  
 QY 410 Tyr--GlyGlyLysAspArgGlyGluAspGluGluSerArgLysArgMetAlaSerGlu 428  
 Db 1324 TATAACAACGGCAAGACCGAGGGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1383  
 QY 429 ValThrAsnAsnLysGlyAsnLeuGluAspArgCysLeuSerCysGlyLysLysAsnPro 448  
 Db 1384 GTTCCCAACAACAGAGAGAGCTGGAAGATGGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1443  
 QY 449 ValSerPheHisProLeuPheGluGlyGlyLeuCysGlnSerCysArgAspArgPheLeu 468  
 Db 1444 GTGCTCTTCCACCTCTCTTGGAGGGGGGCTGTGTGAGATGATGATGATGATGATGATGATGAT 1503  
 QY 469 GluLeuPheTyrMetTyrAspGluAspGlyTyrGlnSerTyrCysThrValCysCysGlu 488  
 Db 1504 GAGCTGTTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1563  
 QY 489 GlyArgGluLeuLeuLeuCysSerAsnThrSerCysArgCysPheCysValGluCys 508  
 Db 1564 GCGGAGAGCTGCTGCTTTCAGACACACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1623  
 QY 509 LeuGluValLeuValGlyAlaGlyThrAlaGluAspAlaLysLeuGlnLysProTrpSer 528  
 Db 1624 CTGAGAGGTGCTGGGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1683  
 QY 529 CysTyrMetCysLeuProGlnArgCysHisGlyValLeuAlaGlyArgLysAspTyrAsn 548  
 Db 1684 TGTACACATGCTCTCCGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1743  
 QY 549 MetArgLeuGlnAspPheThrThrAspProAspLeuGluGlnLysLeuGluProTrpLys 568  
 Db 1744 GTGGGCTCGAGGCTTCTTTCACACAGTACACAGGGGCTT--GAATACAGAACCCCCCAAG 1800  
 QY 569 LeuTyrProAlaIleProAlaAlaLysArgArgProIleArgValLeuSerLeuPheAsp 588  
 Db 1801 CTGTAACCTGCTCATCCCGCAGCCGAGAGCGGCCCATTCAGATCTCTCATTTGTTGAT 1860  
 QY 589 GlyIleAlaThrGlyTyrThrLeuValLeuLysGluLeuGlyIleLysValGluLysTyrIle 608  
 Db 1861 GGCATCGGAGAGCTACCTGCTCTCAAAAGTTGGGCATTAAGATGAGAAAGTACCTC 1920  
 QY 609 AlaSerGluValCysAlaGluSerIleAlaValGlyThrValLysHisGlyGluIle 628  
 Db 1921 GCTTCTGAGAGTGTGAGAGATCTTCTGTTGGAACCGTGAACACGAGAGGAGATATTC 1980  
 QY 629 LysTyrValAsnAspValArgLysIleThrLysLysAsnIleGluGluTrpGlyProPhe 648

Db 1981 AATATCGTAGACGCTGAGAAACATTCACAAAGAAAATTGTAAGATGGGGCCATT 2040  
 QY 649 AspLeuValIleGlyGlySerProCysAsnAspLeuSerAsnValAsnProAlaArgLys 668  
 Db 2041 GACTGGTGAATGGCGGAACCCATGCAACGATCTTCAATGTGAATCCAGCCAGAAA 2100  
 QY 669 GlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPheThrHisLeuAsnTyrThr 688  
 Db 2101 GGCTGTATAGGGTACAGGCCGCTCTTCTTCGAATTTTACCACCTCGTAATTACTCA 2160  
 QY 689 ArgProLysGluGlyAspAsnArgProPhePheThrMetPheGluAsnValAlaAlaMet 708  
 Db 2161 CGCCCAAGAGAGGTGTAGACCGGCCCTTCTTGTGATGTTTGAAATGTTGAAGCATG 2220  
 QY 709 LysValLeuAspLysLysAspIleSerArgPheLeuAlaCysAsnProValMetIleAsp 728  
 Db 2221 AAGTTGGCCACAGAGGAGCATCTTCACGGTTCTCGAAGTGAATCCAGTGAATTTGAT 2280  
 QY 729 AlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMet 748  
 Db 2281 GGCATCAAGTTTCTGCTGCTCAGAGGCCGATATTCTGGGGCACTACCCGGGATG 2340  
 QY 749 AsnArgProValMetAlaSerLysAsnAspLysLeuGluLeuGluAspCysLeuGluPhe 768  
 Db 2341 AACAGGCCCGTGAATGATCAAGAAATGATAACTCAGCTGCAGAGCTGCTTGAATAC 2400  
 QY 769 SerArgThrAlaLysLeuLysLysValGlnThrIleThrThrLysSerAsnSerIleArg 788  
 Db 2401 AATAGGATAGCCAGTTTAAGAAAGTACAGACATTAACCACTGAGTGAATCTGATCAAA 2460  
 QY 789 GlnGlyLysAsnGlnLeuPheProValAlaMetAsnGlyLysAspAspValLeuTrpCys 808  
 Db 2461 CAGGGGAAAACCAACTTTTCCCTGTTGTCATGATGGAAGAAAGATGTTTGTGTCG 2520  
 QY 809 ThrGluLeuGluArgIlePheGlyPheProAlaHisTyrThrAspValSerAsnMetGly 828  
 Db 2521 ACTGAGCTCCAAAGAGATCTTGGCTTTCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2580  
 QY 829 ArgGlyAlaArgGlnLysLeuLeuGluArgSerTrpSerValProValIleArgHisLeu 848  
 Db 2581 CGTGTGCCCCGCAAGAGCTGCTGGGAAGTCTTGAGGCTGCTGTATCTGCACACTTC 2640  
 QY 849 PheAlaProLeuLysAspTyrPheAlaCysGlu 859  
 Db 2641 TTGCGCCCTCGAAGACTACTTTGATGTGA 2673  
 RESULT 14  
 AF331857 4335 bp mRNA linear PRI 02-JAN-2002  
 LOCUS AF331857  
 DEFINITION Homo sapiens DNA cytosine methyltransferase 3 beta (DNMT3B) mRNA,  
 complete cds.  
 ACCESSION AF331857  
 VERSION AF331857.1 GI:18033254  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euleleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 4335)  
 AUTHORS Ni,J., Pradhan,S. and Roberts,R.J.  
 TITLE Cloning, expression and characterization of human DNMT3 genes  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 4335)  
 AUTHORS Ni,J., Pradhan,S. and Roberts,R.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-DEC-2000) New England Biolabs, 32 Tozer Road,  
 Beverly, MA 01915, USA  
 FEATURES  
 source  
 1..4335  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..4335  
 gene



```

/ gene="DNMT3B"
/ 322..2883
/ gene="DNMT3B"
/ codon_start=1
/ product="DNA cytosine methyltransferase 3 beta"
/ protein_id="AL57040.1"
/ db_xref="GI:18033258"
translation="MKGDTRLHNGEEDAGREDSILVNGACSDQSDSPPILEAIRTP
EIKRRSSRLSKREVSSLLSTTQDLGDDGDDGSDTPMVKPRATETATRESFA
VRKNNSVSRERHPRSPTRSGRQNHVDESPVEFPATASLRARATASGTPWS
PPSEYTLIDLTDETHGTPOSSSTPYARLAQSQGGMESFQVEAUSGDCDSBQ
DGKEFGDGLVWGKIKFSGWAMPVMSKATSKRQMSGRWQWFGDGKSEFSAUK
LVALGFSQHFNLAENKLVSYRKAMHALEKVRAGKTFFSPGDSLEQLAPMLE
WAHGGKPTGIEGLKPNNTTPVVKSKVRRAKSKLESRYENKTRRTADDSATSY
CPAPKLTNCYNNKDRGDEQREQMSADVANNKSLEDDGSCGRKNPVSPHPLV
EGGLCOTCDRLPELELFYMYDDGVSCTCCGREGELLCSNTSCRCFCVECLEVLV
GTGTAELKQEPWSCYMLCPQRCGHVLRKKNVRLQAFFTSDTGLYEAPKLYPA
IPARRNPRIKVLISLFDGIATGYLVKELGIVKGYVASECESIATVTEKHEGNIKY
VNDNRNTTKNIIEEMGPFDLVIGSPCNDLSNVNPAKGLYEGTGRLFFEFYHLNTS
RPKEGDDRPFPFMFENVMAMKVKRDI SRPLECNPMIDAIKVAARHARYFNGNLP
GMNRPTVASKNDKLEQDCLEYNRIAKLVQTTTTSKNSIKQGNQLFPVVMNGRED
VLWCTELERIFGPPVHYTVDVSNMGRGAQOKLLGRSMVSVIRHLEAPLKDVFACE"
BASE COUNT 1045 a 1172 c 1167 g 951 t
ORIGIN

Alignment Scores:
Pred. No.: 1,33e-202 Length: 4335
Score: 3759.00 Matches: 717
Percent Similarity: 86.53% Conservative: 55
Best Local Similarity: 82.32% Mismatches: 69
Query Match: 81.90% Indels: 30
DB: 9 Gaps: 10

US-09-720-086-6 (1-859) x AF331857 (1-4335)
Qy 1 MetLysGlyAspSerArgHisLeuAsnGluGluGlyAlaSerGlyTyrGluGluCys 20
Db 322 ATGAAGGGAGACACAGGACATCTCAATGGAGAGAGGAGCGCGCGGAGGAAAGATCG 381
Qy 21 IleIleValAsnGlyAsnPheSerAspGlnSerAspThrLysAspAlaProSerPro 40
Db 382 ATCTCTGTCACACGGGGCTTGCAGCGACAGTCCTCCGAC-----TCGCC 426
Qy 41 ProValLeuGluAlaIleCysThrGluProValCysThrProGluThrArgGlyArgArg 60
Db 427 CCAATCTGGAGGCTATC-----CGCACCCCGAGATCAGAGGCGCGAAGA 471
Qy 61 SerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTyrThrGlnAspMet 80
Db 472 TCAAGCTCGGACTCTCCACAGAGGAGGTGTCCAGTCTGCTAAGCTACACACAGGACTTG 531
Qy 81 ThrGlyAspGlyAspArgAspAspGluValAspAspGlyAsnGlySerAspIle---Leu 99
Db 532 ACAGGCGATGGCGACGG-----GAAGATGGGATGGCTCTGCACACCCAGTC 579
Qy 100 MetProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaVal 119
Db 580 ATGCCAAAGCTCTTCGGGAA-----ACCAGAGCTCGTTCAAGAACCCAGCTGTC 630
Qy 120 ArgThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnAlaSerProArgIle 139
Db 631 CGAACTCGAAATAACAACAGTGTCTCCAGCGCGGAGAGGACAGACCCCTCCCCACGPTCC 690
Qy 140 ThrArgGlyArgGlnGlnArgHisHisValGlnGluTyrProValGluPheProAlaThr 159
Db 691 ACCGAGGCGCGGCGCGCAACCATGTGGACAGTCTCCCGTGGAGTTCCTCCGCTACC 750
Qy 160 ArgSerArgArgArgAlaSerSerSerAlaSerThrProTyrSerSerProAlaSer 179
Db 751 AGGTCCCTGAGACGGCGGCGCAACAGCATCGGACGAGAACGCGCATGTGCGCGTCCCTCCAGC 810
Qy 180 -----ValAspPheMetGluGluVal-----ThrProLysSer 190

```

```

Db 811 TCTTACCTTACCATCGAAGCTTACAGACGACACACAGAGGACACACATGGGAGCCCCAGAGC 870
Qy 191 ValSerThrProSerValAspLeuSerGlnAspGlyAspGlnGluGlyMetAspThrThr 210
Db 871 AGCAGTACCCCTTACGCCCGCTAGCCAGGACAGCCAGCAGCGGGGAGTGGAGTCCCGC 930
Qy 211 GlnValAspAlaGluSerArgAspGlyAspSerThrGluTyrGlnAspAspIleGluPhe 230
Db 931 CAGGTGGAGGACAGACAGTGGAGATTCAGAGTATCAGAGTATCAGGATGGGAAGAGGATTT 990
Qy 231 GlyIleGlyAspLeuValTrpGlyValIleGlyValIleGlyValIleGlyValIleGly 250
Db 991 GGAATAGGGACCTCGTGTGGGGAAGATCAAGGCTTCTCTGTGTGGCCCGCCATGGTG 1050
Qy 251 ValSerTrpLysAlaThrSerLysArgGlnAlaMetProGlyMetArgTrpValGluTrp 270
Db 1051 GTGTCTTGGAGGACCACTTCCAGGCGACAGCTATGTCTGGCATCGGTGGGTCCAGTGG 1110
Qy 271 PheGlyAspGlyLysPheSerGluIleSerAlaAspLysLeuValAlaLeuGlyLeuPhe 290
Db 1111 TTTGGCATGGCAAGTTCTCCAGGCTCTCTGACAGCAAACTGGTGGACATGGGGCTGTTC 1170
Qy 291 SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyr 310
Db 1171 AGCCAGCACTTTAATTTGGCCACCTTCAATTAAGCTCTCTCTATCGAAAGCCATGTAC 1230
Qy 311 HisThrLeuGluLysAlaArgValArgAlaGlyLysThrPheSerSerSerProGlyGlu 330
Db 1231 CATGCTCTGGAGAAAGCTAGGCTGGAGCTGGCAAGACCTTCCCCAGCAGACCTGGAGC 1290
Qy 331 SerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGlyLysProThr 350
Db 1291 TCATTTGGAGGACCACTGAAGCCCATGTTGGAGTGGGCCCCACGGGGCTTCAAGCCACT 1350
Qy 351 GlyIleGluGlyLysProAsnLysGlnProValValAsnLysSerLysValArg 370
Db 1351 GGGATTCAGGGGCTTCAAAACCAACACGCAACAGCTGGTTAATTAAGTCAGAGTGGCT 1410
Qy 371 ArgSerAspSerArgAsnLeuGluProArgArgGluAsnLysSerArgArgThr 390
Db 1411 CDTGCGGAGTGGAAATAGTAATCAAGAAATACAGAACCAAGACTCGAAGACGCA 1470
Qy 391 ThrAsnAspSerAlaLysSerGlu---SerProProLysArgLysLeuLysThrAsnSer 409
Db 1471 GGTGACCACTCAGCCACTCTGACTACTGCCCCCGCACCAAGCGCTCAAGACAAATTCG 1530
Qy 410 Tyr---GlyGlyLysAspArgGlyGluAspGluSerArgGluArgMetAlaSerGlu 428
Db 1531 TATAACAACGCAAGACCGAGGATGAAGATCAGAGCCGAGAACAAATGGCTTCAGAT 1590
Qy 429 ValThrAsnLysGlyAsnLeuGluAspArgCysLeuSerCysGlyLysLysAsnPro 448
Db 1591 GTTGCAACCAAGACGACCTGGAAGATGGCTGTTTGTGTCAGGAGAAACCC 1650
Qy 449 ValSerPheHisProLeuPheGlyGlyLeuCysGlnSerCysArgAspArgPheLeu 468
Db 1651 GTGTCTTCCACCTCTCTTGGGGGGGCTCTGTCAAGATGCCGGATCGCTTCCTT 1710
Qy 469 GluLeuPheTyrMetTyrAspGluAspGlyTyrGlnSerTyrCysThrValCysGlu 488
Db 1711 GAGCTGTTTACATGATGATGACGTGCTATCAGTCTTACTGCTGCTGCTGCTGCTG 1770
Qy 489 GlyArgGluLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCysValGluCys 508
Db 1771 GGGCAGAGCTGTGCTTTCAGAACACAGAGCTGCTGCGGTGTTCTGTGTGGAGTGC 1830
Qy 509 LeuGluValLeuValGlyAlaGlyThrAlaGluAspAlaLysLeuGlnProTyrSer 528
Db 1831 CTGGAGGTGCTGGTGGGACAGACAGCGCGCGGAGCCAGCTTCAGGAGCCCTGGAGC 1890
Qy 529 CysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgArgLysAspThrAsn 548
Db 1891 TGTACATGTCTCTCCGACGCTGTCATGCGCTCTGTCGCGCCCGGAGGACTCGAAC 1950

```



QY 549 MetArgLeuGlnAspPheThrThraSerProAspLeuGlnGluPheGlnProProlys 568  
Db 1951 GTGGCGCTGAGCGCTTCTTACACGATGACACGGGGCTT---GAATATGAAGCCCCCAAG 2007  
QY 569 LeuArgProAlaIleProAlaAlaLeuArgArgProIleAlaGValLeuSerLeuPheAsp 588  
Db 2008 CTGTAACCTCGCATCTCCGCGAGCGGCCCATTCAGAGTCCTGTCTCATTTGTGAT 2067  
QY 589 GlyIleAlaThrGlyIleLeuValLeuLeuGlnGluGlyIleValGluLeuTyrIle 608  
Db 2068 GGCAATCCGACAGAGCTACTGCTCTCAAGAGTTGGGCAATAAGTGAAGTACGTC 2127  
QY 609 AlaSerGlnValCysAlaGlnSerIleAlaValGlyThrValLeuSerGlnGlyGlnIle 628  
Db 2128 GCTTCTGAAGTGTGAGAGACTGCTGTTGGAACCGTGAAGACAGAGGGGAATTC 2187  
QY 29 LysArgValAsnAspValArgValIleThrLysLysAsnIleGlnGluTTPGlyProPhe 648  
Db 2188 AATATCGTGAACGAGTGAAGACATCAAAAGAAAATATGAAAGATGGGGCCCATTT 2247  
QY 649 AspLeuValIleGlyGlySerProCysAsnAspLeuSerAsnValAsnProAlaArgLys 668  
Db 2248 GACTTGGTGAATGGCGAGAGCCCATGCAACGATCTCTCAATGTAATCCAGCCAGAAA 2307  
QY 669 GlyLeuTyrGlnGlyIleThrGlyArgLeuPhePheGluPheTyrHisLeuLeuAsnTyrThr 688  
Db 2308 GGCTGTATGAGGGGTACAGCGCGGCTTCTTGAAATTTTCAACCTGCTGAATTAACA 2367  
QY 689 ArgProLysGlnGlyAspAsnArgProPheThrPheMetPheGluAsnValAlaIleMet 708  
Db 2368 CGCCCAAGAGGGGTGAAGACCGCGCTTCTTGAGATTGAAAGTGTAGCCCATG 2427  
QY 709 LysValAsnAspLysLysAspIleSerArgPheLeuAlaCysAsnProValMetIleAsp 728  
Db 2428 AAGTTGGCGCAAGAGGGAATCTCACGCTTCTGAGGTGAATCCAGTATGATTGAT 2487  
QY 729 AlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMet 748  
Db 2488 GCCATCAAAAGTTTCTGCTGCTCACAGGGCCCGATCTTCTGGGGAACTACCCGGGATG 2547  
QY 749 AsnArgProValMetAlaSerLysAsnAspLysLeuGlnAspCysLeuGluPhe 768  
Db 2548 AACAGGCCCGGATGACATCAAGAGATGATTAACCTCGAGCGAGCACTGCTGGAATAC 2607  
QY 769 SerArgThrAlaLysLeuLysValGlnThrIleThrLysSerAsnSerIleArg 788  
Db 2608 AATAGGATAGCCAGATTAAAGAAAGTACAGCAATAACCAACCAAGTCAACTCGATCAAA 2667  
QY 789 GlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysAspAspValLeuTyrCys 808  
Db 2668 CAGGGGAAAAACAACCTTTCCGTGTTGTCAATGCAAGAAAGATGTTTGTGGTGC 2727  
QY 809 ThrGluLeuGlnAlaIlePheGlyPheProAlaHisTyrThrAspValSerAsnMetGly 828  
Db 2728 ACTAGAGTCGAAAGAGATCTTGGCTTCTGTGACATCAACAGAGCTGTCCAAACATGGGC 2787  
QY 829 ArgGlyAlaArgGlnLysLeuLeuGlnLysArgSerTyrSerValProValIleArgHisLeu 848  
Db 2788 CGTGGTGGCCCGCAGAGCTCTGGGAAAGTCTCGAGAGCGTGTATCCGACACCTC 2847  
QY 849 PheAlaProLeuLysAspTyrPheAlaCysGln 859  
Db 2848 TTGCGCCCTCGAAGGACTACTTGGCATGTGAA 2880  
RESULT 15  
LOCUS AF176228 4267 bp mRNA linear PRI 26-OCT-1999  
DEFINITION Homo sapiens DNA cytosine-5 methyltransferase 3B (DNMT3B) mRNA,  
complete cds.  
ACCESSION AF176228  
VERSION AF176228.1 GI:6118091

SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Xu, G.-L., Bestor, T.H., Bourc'his, D., Hsieh, C.-L., Tommerup, N.,  
Bugge, M., Hulten, M., Qu, X., Russo, J.J., and Viegas-Pequignot, E.,  
Chromosome instability and immunodeficiency syndrome caused by  
mutations in a DNA methyltransferase gene  
Nature (1999) In press  
2 (bases 1 to 4267)  
XU, G.-L. and BESTOR, T.H.  
Direct Submission  
Submitted (06-AUG-1999) Genetics and Development, Columbia  
University, 701 West 168 St., New York, NY 10032, USA  
LOCATION/Qualifiers  
1..4267  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/map="20q11-q13"  
1..4267  
/gene="DNMT3B"  
265..2802  
/gene="DNMT3B"  
/note="mutated in human chromosome instability and  
immunodeficiency disease known as ICF syndrome; contains  
alternative 5' exon 1b"  
/codon\_start=1  
/product="DNA cytosine-5 methyltransferase 3B"  
/protein\_id="AA04015.1"  
/db\_xref="GI:6118092"  
/translation="MEPSPEPSELSMKGDTFRLNGEDPAGREDSILVNGACSDSS  
DSPRLAETRRPEIRGERSSPRLSEKREVALSYDGLDGGDGGEDSDPVMKL  
FRRTFRSSPAVTRTNNSVSSREHRRSPSTRQGRNRHDSPEVPRATRSIR  
RATAGTPEPSPSSYLITDITDIEDTHGTPQSSSTPYARLAQDSQGGMESPOVE  
ADSGDSSSEYDQKEFGIDLVWGIKQPSVMPAVVSWKATSKQASGNWVQVE  
GGKFESEVADKLVALGLFSOHNLATFNKLVSRYKAMHALAKARVAKTPSSFG  
DSLEDQKPMLEWAGHGFPTGIEGLKPNNTOPENRTRRRTADSDATSDYCAPKSLK  
TNCVNNQKRGDEDSREOMASPVANNKSLSDGCLSCRKNPVSPHPIPEGGLCOTC  
RDRFLFLPTMYDDDGOSYCTYCCERELLCSNTSCRCFCYCECLFLVGTATAEA  
KQEPNSCYMCLPQRCRGVLRKRDNNVLAQAFSTDTGLEEAPLPAIPARRRP  
IRVLSLPDGIANGYLVLEKLGIKVGYVASEVEESIAVGTVHGBGNIKYVNDVNT  
KRIIEBGFDFDVIIGSPNDLSNVPAKGLVEGTGRIFFEYHLLNYSRPREGDR  
PFPMFENVVAMKVGDKRDISPRLCNPVVIAIKYSAAHARVPMGNLPGMNPVIA  
SKNDKLELDODCEVNRILAKLYQVTTTNSIKQKQKQLFVYVWAKGKEDVLMCTBLE  
RIFGFPVHTDVSNMGRARQKULGRSMVSPVIRHLPAIPUKDTFACE"  
BASE COUNT 1069 a 1098 c 1147 g 953 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 7, 01e-198 Length: 4267  
Score: 3675.00 Matches: 703  
Percent Similarity: 86.80% Conservative: 53  
Best Local Similarity: 80.71% Mismatches: 65  
Query Match: 80.07% Indels: 50  
DB: 9 Gaps: 11  
US-09-720-086-6 (1-859) x AF176228 (1-4267)  
QY 1 MetCysGlyAspSerArgHisLeuAsnGlnGluGlnLysAlaSerGlyTyrGlnGluCys 20  
Db 301 ATGAAGGAGAGACACGACGATCTCAATGAGAGAGAGACCGCGGAGAGAGACTCG 360  
QY 21 IleIleValAsnGlyAsnAspPheSerAspGlnSerSerAspThrLysAspAlaProSerPro 40  
Db 361 ATCTCTCGTCAACGGGGGCTCGACGACGACGATCTCCACAC-----TTCGCC 405  
QY 41 ProValLeuGlnAlaIleCysThrGluProValCysThrProGluThrArgGlyArgArg 60  
Db 406 CCAATCTCGAGGGCTATC-----CGCACCCCGGAGATCAAGCGCGGAGAGA 450  
QY 61 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTyrThrGlnAspMet 80

```
Db 451 TCAGCTCCGACTCTCCAGAGGGAGGTGTCCAGTCTGCTAAGCTACACAGGACTTG 1510
Qy 81 ThrGlyAspGlyAspArgAspAspGluValAspAspGlyAsnGlySerAspIle---Leu 99
Db 511 ACAGGCGATGGCGAGCGG-----GAGATGGGGATGGCTCTGACACCCAGTC 558
Qy 100 MetProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaVal 119
Db 559 ATGCCAAAGCTCTCCCGGAA-----ACCAGAGCTCGTTTCAGAAAGCCAGCTGTC 609
Qy 120 ArgThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArgAlaSerProArgIle 139
Db 610 CGAACTCGAAATAACAACAGTGTCTCCAGCCGGGAGAGCACAGCCCTCCCCACGTTCC 669
Qy 140 ThrArgGlyValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 159
Db 670 ACCGAGGCGCGGCGGCAACCATGTGGACGAGTCCCGTGGAGTTCCCGGTACC 729
Qy 160 ArgSerArgArgArgAlaSerSerSerAlaSerThrProTrpSerSerProAlaSer 179
Db 730 AGGTCCCTCAGACGCGCGGCAACAGCATCGCAGGAAACGCCATGGCCCTCCCTCCAGC 789
Qy 180 -----ValAspPheMetGluGluVal-----ThrProLysSer 190
Db 790 TCTTACCTTACCATCGACTCCACAGACGACACAGAGGACACACATGGGAGCGCCCGAGC 849
Qy 191 ValSerThrProSerValAspLeuSerGlnAspGlyAspGlnGlnGlnGlnGlnGlnGln 210
Db 850 AGCATGACCCCTCAGCCCGCTAGCCAGGACGACGACGAGCGGGGCGATGGAGTCCCGC 909
Qy 211 GlnValAspAlaGlnSerArgAspGlyAspSerThrGluThrGlnAspAspLysGluPhe 230
Db 910 CAGGTGGAGGACAGACAGTGGAGTGGAGACAGTTCAGAGTATCAGATGGGAGAGTTT 969
Qy 231 GlyIleGlyAspLeuValTrpGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGly 250
Db 970 GGAATAGGGAGCTCGTGGGGAAAGATCAGGGCTTCTCTGGTGGCCCGCATGGTG 1029
Qy 251 ValSerTrpLysAlaThrSerLysArgGlnAlaMetProLysMetArgTrpValGlnTrp 270
Db 1030 GTGCTTGGAGGACACTCCAGCGACAGCTATGTCGTGGATCGGTGGGTCCAGTGG 1089
Qy 271 PheGlyAspGlyLysPheSerGluIleSerAlaAspLysLeuValAlaLeuGlyLysPhe 290
Db 1090 TTTGGCGATGGCAAGTTCTCCGAGGTCTCTGCAGACAAACTCGTGGCACTGGGGCTGTC 1149
Qy 291 SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTrpArgLysAlaMetTrp 310
Db 1150 AGCCAGCACTTTAAATTGGCCACCTTCAATAAGCTCGTCTCTATCGAAAAAGCCATGAC 1209
Qy 311 HisThrLeuGluLysAlaArgValArgAlaGlyLysThrPheSerSerSerProGlyGlu 330
Db 1210 CATGCTCTGGAGAAAGCTAGGTGGAGCTGGCAAGACCTTCCCCAGCAGCCCTGGAGAC 1269
Qy 331 SerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGlyPheLysProThr 350
Db 1270 TCATTGGAGGACCAAGTGAAGCCCTGTTGGAGTGGGCCCCACCGGGGGCTTCAAGCCACT 1329
Qy 351 GlyIleGluGlyLysProAsnLysLysGlnProValValAsnLysSerLysValArg 370
Db 1330 GGGATCGAGGCTTCAAAACCCAAACACCGCAACA----- 1365
Qy 371 ArgSerAspSerArgAsnLeuGluProArgArgGluAsnLysSerArgArgThr 390
Db 1366 -----GAGAACAGACTCGAAGAGCCACA 1389
Qy 391 ThrAsnAspSerAlaAlaSerGlu---SerProProLysArgLeuLysThrAsnSer 409
Db 1390 GCTACACACTCAGCCACTCTGACTACTGCGCCCGACCCAGGCGCTTCAAGCAANTGCG 1449
Qy 410 Tyr---GlyGlyLysAspArgGlyLysAspGluLysSerArgGluArgMetAlaSerGlu 428
Db 428
```

```

Qy 789 GlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysAspAspValLeuTrpCys 808
    |||||
Db 2587 CAGGGGAAAAACCAACTTTCTCTGTGTCAATGCAAGAAAGATGTTGTGGTGC 2646
    |||||
Qy 809 ThrGlnLeuGlnIuArgGlyPheGlyPheProAlaHisTyrThrAspValSerAsnMetGly 828
    |||||
Db 2647 ACTGAGCTCGAAAGGATCTTTGGCTTCTGTGCACTACACAGAGGTGCCACATGGGC 2706
    |||||
Qy 829 ArgGlyAlaArgGlnLysLeuLeuGlyArgSerTrpSerValProValIleArgHisLeu 848
    |||||
Db 2707 CGTGTGCGCCGCCAGAGCTGTGTGGAAAGTCTGTGAGCGTGTGCTGTATCCGACACCTC 2766
    |||||
Qy 849 PheAlaProLeuLysAspTyrPheAlaCysGlu 859
    |||||
Db 2767 TTCGCCCTCTGAAGACTACTTGTGATGTGAA 2799
    |||||

```

Search completed: November 22, 2002, 10:52:58  
 Job time : 4989.5 secs

**THIS PAGE BLANK (USPTO)**



**THIS PAGE BLANK (USP)**

AF156488  
LOCUS AF156488 4145 bp mRNA linear PRI 05-SEP-1999  
DEFINITION Homo sapiens DNA cytosine-5 methyltransferase 3 beta 1 (DNMT3B)  
ACCESSION AF156488  
VERSION AF156488.1 GI:5823167  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 4145)  
Xie,S., Wang,Z., Okano,M., Nogami,M., Li,Y., He,W.W., Okumura,K.  
and Li,E.  
Cloning, expression and chromosome locations of the human DNMT3  
gene family  
Gene 236 (1), 87-95 (1999)  
JOURNAL  
LINE 99365304  
MED 10433969  
2 (bases 1 to 4145)  
Xie,S. and Li,E.  
Direct Submission  
Submitted (03-JUN-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA  
JOURNAL  
TITLE  
AUTHORS  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
1. 4145  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/map="between D20S182 and D20S106; 39,9-50.2 cM"  
1. 4145  
/gene="DNMT3B"  
115..2676  
/note="DNMT3B"  
/note="de novo DNA methyltransferase"  
/codon\_start=1  
/product="DNA cytosine-5 methyltransferase 3 beta 1"  
/protein\_id="AAD53063.1"  
/db\_xref="GI:5823168"  
/translation="MKGDTRHNGEDAGREDSDILVNGACSDQSSDSEPILEAIRTP  
ETRRGRSSRSKRREVSLLSTYDOLTGDDGDDGSDTPVMPKLFRTTRSESPA  
VTRNNNSVSSRRHPPSPSTRGROGRHVDSPPEPATSLRRATASAGTPWPS  
PPSYLTIDLTDDTDTGTPSSSTPYRLADDSQGGMESPOVADGDSSEYO  
DKDEFGIDLVNFKIKGFSWPMAMVSMATSKRQMSGRKRWQMGDKFSEVSDK  
LVALGFSQFNPLATFNKLVSYRKAMVHLEKRVAGATFPSSPEDSDJDKPMLE  
MAHGKPKPIGELKPNTPOPVYKSKVRAGSKLESRYENKTRRRADDSATSDY  
CPAPKLTKNVNGKRDGDDSDGSDVANNKSLSDGCLSGRKPVSFHLF  
EGGLCOTCHDRFLELFYMDDDGSDYCTVCGRELLCSNCSRCRCEVECLVLY  
GGLTAELAKIQEPSCYMCUPORCHVLRKRMNRLAFTPSDGLYEAKLYPA  
IPARRRPTRVLFDGIATGLVLELGIKVKRYASEVCESSIAVGTYKHEGNTKY  
VNDVRNTTKNIEMGPFDLVIGSPDNDSNVPAPKGLYBGTGLFFPFTHLNLS  
RPEKDDRPFFMENVAMKVDKRDISRFLCNPMYMDIVSAHRRARYMGNL  
GMARPVYASKNDLELDDCLVNRIRAKLVKQVITTTSSNISKQKDLFPVNWGKD  
VLMCTELERIFGSPVHYTDVSNRGRAROKLGRSMSPVIRHLFALVDYFACE"

BASE COUNT 1041 a 1083 c 1096 g 925 t  
ORIGIN  
Alignment Scores: 5.02e-234 Length: 4145  
Pre: 4566.00 Matches: 853  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 100.00% Gaps: 0  
DB: 9  
US-09-720-086-8 (1-853) x AF156488 (1-4145)  
QY 1 MetLysGIAyPThrArgHisLeuAnnglygluGluAepAlGlyGlyArgGluAspSer 20  
DB 115 ATGAAGGAGACACACGAGCATCTCAATGAGAGAGAGACCGCGCGGAGGAGAGACTCG 174  
21 TLeuValaAnnglyAlaCySerAepGlnSerSerAepSerProProlleuGluAla 40

DB 175 ATCTCTGTCACAGGGGCTTCGACACACAGTCTCTCGACCTCGCCCAATCTCGAGGCT 114  
QY 41 TLeArgThrProGluIleArgGlyArgArgSerSerArgLeuSerLysArgGluVal 60  
DB 235 ATCCGACACCCCGAGATTCAGAGCCGCAAGATCAAGCTCCGACATCTCCAGAGAGAGTG 294  
QY 61 SerSerLeuSerLysThrGlnApleuThrGlyAspGlyAspGlyGluAspGlyAsp 80  
DB 295 TCCAGTCTGCTAAGCTACACACAGAGACTTGACAGGCGATGGCAAGGGAAGATGGGAT 354  
QY 81 GlySerAepThrProValMetProLysLeuPheArgGluThrArgThrArgSerGluSer 100  
DB 355 GGCTTCACACCCAGTCATGCTCAAGACTCTCCGGGAAACAGGACTCGTTCAGAAAGC 414  
QY 101 ProAlaValArgThrArgAmaAmaSerValSerSerArgGluValGlnArgProSer 120  
DB 415 CCAGCTGTCGAACTCGAAATTAACAACAGTGTCTCCGCGGGAGAGACAGGCTTCC 474  
QY 121 ProArgSerThrArgGlyArgGlnGlyArgAsnHisValAspGluSerProValGluPhe 140  
DB 475 CCAGCTTCACACCCAGAGCGCGGAGCGGCAACATGTGACAGTCTCCCTGGATTC 534  
QY 141 ProAlaThrArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProTrpProSer 160  
DB 535 CCGGCTACAGAGTCTCTAGACCGCGGCAACAGATCGCAGAGAACGCCATGGCTTCC 594  
QY 161 ProProSerSerLysLeuThrIleAspLeuThrAspAspThrGluAspThrHisGlyThr 180  
DB 595 CCTCCAGCTCTTACCTTACCATCGACCTCAGACGACGACAGACAGACACACATGGGAC 654  
QY 181 ProGlnSerSerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMet 200  
DB 655 CCCAGAGACAGACGATACCCCTCAGCCGCTAGCCGACGAGACGACGACGAGGGGCGATG 714  
QY 201 GluSerProGlnValGluAlaAspSerGlyAspGlyAspSerSerGluTyrGlnAspGly 220  
DB 715 GAGTCCCGCAGCTGAGGACACACAGTGGAGATGGAGATTCAGATCAGATCAGATGCG 774  
QY 221 LysGluPheGlyIleGlyAspLeuValTyrGlyLysIleLysGlyPheSerTrpTrpPro 240  
DB 775 AAGGATTTTGAAATAGGAGACCTCGTGTGGGAAAGATCAAGGCTTCTCTGTGGTCCC 834  
QY 241 AlaMetValAlaSerTrpLysAlaThrSerLysArgGlnAlaMetSerGlyMetArgTrp 260  
DB 835 GCCATGTGCTGCTTGAAGGCGACCTCCAGGCTCTCTGACATCGGCTG 894  
QY 261 ValGlnTrpPheGlyAspGlyLysPheSerGluValSerAlaAspLysLeuValAlaLeu 280  
DB 895 GTCCAGTGTGTGGCGATGGCAAGTCTCCGAGTCTCTGACAACTGGTGGCACTG 954  
QY 281 GlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerLysArgLys 300  
DB 955 GGGGTGTTACGCGACGACTTAATTGGCCACCTTCATTAAGCTCGTCTCTAAGAAA 1014  
QY 301 AlaMetTyrHisAlaLeuGluLysValAlaArgValArgAlaGlySerThrPheProSer 320  
DB 1015 GCCATGTACCATGCTCTGCGAAGAAAGCTAGGCTGAGCTGCGAAGACTTCCACAGAGC 1074  
QY 321 ProGlyAspSerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGlyPhe 340  
DB 1075 CTTGAGACTCATTTGAGAGACGACGACTGAAGCCCATGTTGGAGTGGCCACGAGGGGCTTC 1134  
QY 341 LysProThrGlyIleGluGlyLeuLysProAsnAsnThrGlnProValAlaAsnLysSer 360  
DB 1135 AAGCCACTGGAGTCGAGGCGCTCAAAACCAACACGCAACGAGTGTAAATTAAGTCG 1194  
QY 361 LysValAlaArgAlaGlySerArgLysLeuLysLeuSerArgLysArgLysLeuThrArg 380  
DB 1195 AAGGTGCTCTGTCAGAGGAGATGAATTAATTAATTAATTAATTAATTAATTAATTA 1254  
QY 381 ArgArgThrAlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLeuLys 400  
DB 1255 AGACGACAGCTGACGATCAGCACTCAGCACTCAGCACTCAGCACTCAGCACTCAGCACT 1314

**THIS PAGE BLANK (USPTO)**



QY 401 ThrAenCyTyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluGlnMet 420  
DB 1315 ACAAAATTCCTATTAACAACGGCAAGACCGAGGGGATGAAGATCAAGAGCGCAAGCAAAATG 1374  
QY 421 AlaSerAspValAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerArgGlyArg 440  
DB 1375 GCTTACAGATGTTGGCAACAACAGACAGCTGGAAAGTGGCTGTTGCTTGTGGCAGG 1434  
QY 441 LysAsnProValSerPheHisProLeuPheGluGlyGlyLeuGluGlnThrCysAsnArg 460  
DB 1435 AAAAACCCTGCTCTTCCACACCTCTCTTGAAGGGGGGCTGTGTACAGACATCGCCGAGT 1494  
QY 461 ArgPheLeuGluLeuPheThrMetTyrAspAspAspGlyTyrGlnSerTyrCysThrVal 480  
DB 1495 CGCTCCCTTGAGCTCTTTTACATGTATGATGATGATGCTATCAGCTCTACACGCACTGTG 1554  
QY 481 CysCysGluGlyArgGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 500  
DB 1555 TGCTGCGAGGGCGGAGAGCTGCTGCTTGGCAGCAACAGCTGCTGCGGAGTCTTCTT 1614  
QY ValGluCysLeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnGlu 520  
DB GTGAGTCTGCTGAGAGTCTGTGTTGGCACAGCACAGCGCGCAAGCTTCAGAG 1674  
QY 521 ProTyrPserCysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgGlyLys 540  
DB 1675 CCTTGACCTGTACATGTGTCTCCCGCAGCGCTGTATGCGCTCTGCGCGCGCAAG 1734  
QY 541 AspTyrAsnValArgLeuGlnAlaPhePheThrSerAspThrGlyLeuGluTyrGluAla 560  
DB 1735 GACTGGAAGCTGCGCTGCGAGCTTCTTCCACAGTACAGCGGGCTTGAAATACGAAGCC 1794  
QY 561 ProLysLeuTyrProAlaLeuProAlaAlaArgArgArgProLeuArgValLeuSerLeu 580  
DB 1795 CCCAAGCTGTACCTGCGCACTTCCCGAGCGCGCAAGCGGCCCATTCGATCGATTCG 1854  
QY 581 PheAspGlyTyrAlaThrGlyTyrLeuValLeuLysGlyLeuGlyTyrLysValGlyLys 600  
DB 1855 TTTGATGGCATGCGCAAGCTGACCTAGCTCTCAAGAGTTGGCATTAAGGTAGGAAAG 1914  
QY 601 TyrValAlaSerGluValCysGluGluSerIleAlaValGlyThrValLysHisGluGly 620  
DB 1915 TAGCTGCTTCTGAAGTGTGTAGAGAGTCTGATTCGTTGAAACCGTGAAGCAACGAGGG 1974  
QY 621 AsnIleLysTyrValAsnAspValArgAsnIleThrLysLysAsnIleGluGluTyrGly 640  
DB 1975 AATATCAAAATAGTGAACGACGAGGAAATCACAAAGAAAATTTGAAGATGGGGC 2034  
QY 641 ProPheAspLeuValIleGlyLysSerProCysAsnAspLeuSerAsnValAsnProAla 660  
DB 2034 CCATTTGACTTGTGATTTGGCGAAGCCCATGCAAGATCTCTCAATGTGAATCCAGCC 2094  
QY 661 ArgLysGlyLeuTyrGluGlyThrGlyArgLeuPheGluPheThrHisLeuLeuAsn 680  
DB 2095 AGCAAAAGCTGTATGAGGCTACAGGCGGCTCTTCTTCAATTTTACCACTGCTGAAT 2154  
QY 681 TyrSerArgProLysGluGlyAspAspArgProPhePheThrPheGluAsnValVal 700  
DB 2155 TACTCAGCGCCCAAGAGAGGTATGACCGGCGCTCTTCTGATGTTTGAAGATGTTGTA 2214  
QY 701 AlaMetLysValGlyAspLysArgAspIleSerArgPheLeuGluCysAsnProValMet 720  
DB 2215 GCCATGAGGTTGGCGAACAAGAGGACATCTCACGGTTCCTGGAGTGAATCCAGTGATG 2274  
QY 721 IleAspAlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuPro 740  
DB 2275 ATTTGATGCGCATTAAGTTTCTCTCTCTCACAGGGCCCGATACTTCTGGGGCAACTACCC 2334  
QY 741 GlyMetAsnArgProValIleAlaSerLysAsnAspLysLeuGluLeuGlnAspCysLeu 760  
DB 2335 GCGATGAACAGCGCCCTGTATGATCAAAAGATGATTAATCTGAGCTGCAAGACTGCTTG 2394

QY 761 GluTyrAsnArgIleAlaLysLeuLysLysValGlnThrIleThrThrLysSerAsnSer 780  
DB 2395 GAATTAACAATAGGATAGCCAAAGTTAAAGAAAGTACAGACAAATTAACCAAGTCGAACTCG 2454  
QY 781 IleLysGlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysGluAspValLeu 800  
DB 2455 ATCAACACAGGGGAAAAACCACTTTTCCCTGTTGTGATGAAATGCAAGAAAGATGTTTG 2514  
QY 801 TyrCysThrGluLeuGluArgIlePheGlyPheProValHisTyrThrAspValSerAsn 820  
DB 2515 TGCTGACTGAGCTGGAAGAAGATCTTGGCTTTCCTGTGACTACACAGACGTGTCCAAAC 2574  
QY 821 MetGlyArgGlyAlaArgGlnLysLeuLeuGluLysArgSerTyrPserValProValIleArg 840  
DB 2575 ATGGGCGGTGTGCTGCGCAAGAGCTGCTGGGAAGGCTCTGGAGCGCTGCTGTATCCGA 2634  
QY 841 HisLeuPheAlaProLeuLysAspTyrPheAlaCysGlu 853  
DB 2635 CACCTCTTGGCCCTCTGAAGACACTTTCATGTGA 2673

RESULT 2  
AF331857  
LOCUS  
DEFINITION Homo sapiens

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 22, 2002, 03:58:32 ; Search time 4860.31 Seconds  
(without alignments)  
5107.633 Million cell updates/sec

Title: US-09-720-086-8  
Perfect score: 4566  
Sequence: 1 MKGDTRHNGEDDAGREDS.....MSVPVIRHLPAPLKDYFACE 853

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 205640 segs, 14551402876 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlp  
-O=/cogn1/USPTO.spool/US09720086/runat.18112002.092832.22195/app.query.fasta\_1.4252  
-DB=GenBdl -QFMT=fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09720086 @CGN 1 11308 @runat.18112002.092832.22195 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THRAD=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: gb\_ba:\*  
2: gb\_hg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_ov:\*  
7: gb\_ov:\*  
8: gb\_ov:\*  
9: gb\_ov:\*  
10: gb\_ov:\*  
11: gb\_ov:\*  
12: gb\_ov:\*  
13: gb\_ov:\*  
14: gb\_ov:\*  
15: gb\_ov:\*  
16: gb\_ov:\*  
17: gb\_ov:\*  
18: gb\_ov:\*  
19: gb\_ov:\*  
20: gb\_ov:\*  
21: gb\_ov:\*  
22: gb\_ov:\*  
23: gb\_ov:\*  
24: gb\_ov:\*  
25: gb\_ov:\*  
26: gb\_ov:\*  
27: gb\_ov:\*  
28: gb\_ov:\*

29: em\_vi:\*  
30: em\_hg\_hum:\*  
31: em\_hg\_inv:\*  
32: em\_hg\_other:\*  
33: em\_hg\_mus:\*  
34: em\_hg\_pln:\*  
35: em\_hg\_rtd:\*  
36: em\_hg\_mam:\*  
37: em\_hg\_vit:\*  
38: em\_sy:\*  
39: em\_hg\_hum:\*  
40: em\_hg\_mus:\*  
41: em\_hg\_other:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4566	100.0	4145	9 AF156488	AF156488 Homo sapi
2	4566	100.0	4335	9 AF331857	AF331857 Homo sapi
3	4450	97.5	4267	9 AF176228	AF176228 Homo sapi
4	4083.5	89.4	3897	9 AF156487	AF156487 Homo sapi
5	3759	82.3	4195	10 AF068626	AF068626 Mus muscu
6	3748	82.1	4223	10 AF151973	AF151973 Mus muscu
7	3748	82.1	4338	10 AF151969	AF151969 Mus muscu
8	3675	80.5	4135	10 AF068627	AF068627 Mus muscu
9	3664	80.2	4163	10 AF151974	AF151974 Mus muscu
10	3664	80.2	4278	10 AF151970	AF151970 Mus muscu
11	3415.5	74.8	4006	10 AY078427	AY078427 Mus muscu
12	3404.5	74.6	4034	10 AF151975	AF151975 Mus muscu
13	3404.5	74.6	4149	10 AF151972	AF151972 Mus muscu
14	3331.5	73.0	3946	10 AF068628	AF068628 Mus muscu
15	3320.5	72.7	3934	10 AF151976	AF151976 Mus muscu
16	3320.5	72.7	4089	10 AF151971	AF151971 Mus muscu
17	2767.5	60.6	3017	9 AK001191	AK001191 Homo sapi
18	2175.5	47.6	4094	10 BC007466	BC007466 Mus muscu
19	2170.5	47.5	3005	9 AF067972	AF067972 Homo sapi
20	2170.5	47.5	4192	10 AF068625	AF068625 Mus muscu
21	2170.5	47.5	4258	9 AF331856	AF331856 Homo sapi
22	1834.5	40.2	2848	5 AF135438	AF135438 Danio rer
23	1740	38.1	2191	9 AK025230	AK025230 Homo sapi
24	1414	31.0	17697	2 AC112586	AC112586 Rattus no
25	1313	28.8	123936	2 AC120824	AC120824 Rattus no
26	1240	27.2	1758	2 BC018214	BC018214 Homo sapi
27	1183.5	25.9	167568	2 AC111734	AC111734 Rattus no
28	1178	25.8	2127	9 AF129267	AF129267 Homo sapi
29	1161.5	25.4	16851	9 AC009474	AC009474 Homo sapi
30	1020.5	22.3	2057	9 AF129268	AF129268 Homo sapi
31	926	20.3	2008	9 AF129269	AF129269 Homo sapi
32	903	19.8	2077	6 AR129189	AR129189 Sequence
33	889.5	19.5	244329	2 AC107644	AC107644 Mus muscu
34	806.5	17.7	110000	2 AL354832.0	AL354832 Homo sapi
35	806.5	17.7	118899	2 HS108597	HS108597 Human DNA
36	751	16.4	225045	2 AL833803	AL833803 Mus muscu
37	751	16.4	249245	2 AC122356	AC122356 Mus muscu
38	671.5	14.7	1397	9 BC002560	BC002560 Homo sapi
39	669	14.7	1705	9 AF194032	AF194032 Homo sapi
40	661.5	14.5	1589	10 AF220524	AF220524 Mus muscu
41	661.5	14.5	1645	10 MM040467	MM040467 Mus muscu
42	646	14.1	188936	2 AC128509	AC128509 Rattus no
43	615	13.5	210269	2 AC116459	AC116459 Mus muscu
44	608.5	13.3	119630	2 AC112040	AC112040 Rattus no
45	520	11.4	153533	2 AC095199	AC095199 Rattus no

RESULT 1

## ALIGNMENTS



QY 401 ThrAnCysTyrAsnAsnGlyLysAspArgGlyAspGlnSerArgGlnGlnMet 420  
 Db 1315 ACAATTGCTATAACAACGCGAAGACCGAGGGATGATGATCAAGCCGAGCAAAATG 1374  
 QY 421 AlAserAspValAlaAsnAsnLysSerSerLeuGlnAspGlyCysLeuSerCysGlyValArg 440  
 Db 1375 GCTTCAGATGTTGGCAACMAAGAGAGCTGGAAGATGGCTTTGCTTGGCGAGG 1434  
 QY 441 LysAsnProValSerPheHisProLeuPheGlnGlyLeuCysGlnThrCysArgAsp 460  
 Db 1435 AAAAACCCTGGCTCTTCAACCTCTTTGAGGGGGGGCTCTGTCAACATGCCGGAT 1494  
 QY 461 ArgPheLeuGlnLeuPheTyrMetTyrAspAspArgGlyTyrGlnSerTyrCysThrVal 480  
 Db 1495 CGCTTCCTTGAGCTGTTTACATGATGATGATGATGATGATGATGATGATGATG 1554  
 QY 481 CysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 500  
 Db 1555 TGGTGGCAGAGGCGCAGAGAGCTGCTTGGCAGCAACAGAGCTGCTGCTGCTGCT 1614  
 QY 501 ValGlnCysLeuGlnGlnValLeuValGlyThrGlyThrAlaAlaGlnAlaLysLeuGlnGln 520  
 Db 1615 GTGAGAGTCCCTGGAGAGTCTGCTGGGACAGGACAGGCGCGAGCCAGACTTCAGAG 1674  
 QY 521 ProTyrSerCysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgArgLys 540  
 Db 1675 CCCTGGAGAGCTGCTACATGCTGCTCCCGCAGCGCTGTCACTGGCGCTGCTGCGG 1734  
 QY 541 AspTyrAsnValAlaGlnGlnAlaPhePheThrSerAspThrGlyLeuGlnTyrGlnAla 560  
 Db 1735 GACTGGAACGCGCGCTGAGCGCTTCTTCAACAGTGCACAGGGGCTTGAAATACGAA 1794  
 QY 561 ProLysLeuTyrProAlaAlaProAlaAlaArgArgArgProLysValLeuSerLeu 580  
 Db 1795 CCAAGCTGTAACCTGCTGCTCCGACAGCCGAGAGCGGCCCATTCAGCTCTGCTCAT 1854  
 QY 581 PheAspGlyAlaAlaThrGlyTyrLeuValLeuLysGlnGlnGlnGlnGlnGlnGln 600  
 Db 1855 TTTGATGGCATCGGACAGCGCTACTACTCTCAAGATGGTGGGATTAAGTGAAG 1914  
 QY 601 TyrValAlaSerGlnValCysGlnGlnSerLeuAlaValGlyThrValLysHisGlnGly 620  
 Db 1915 TACCTGCTTGTGAAGCTGTGAGAGATCCATTGCTTGTGAACCGTGAACACAGG 1974  
 QY 621 AsnLeuTyrValAlaAsnAspValArgAsnLeuThrLysAsnGlnGlnGlnGlnGln 640  
 Db 1975 AATATCAAAATGCGTGAAGAGCTGAGCAATCAACAAAGAAATATTGAAGAAATGG 2034  
 QY 641 ProPheAspLeuValAlaGlyLysProCysAsnAspLeuSerAsnValAsnProAla 660  
 Db 2035 CCATTTGACTTGGATGATGGGGAAGCCCATGCAACGATCTCTCAAAATGATTCAGCC 2094  
 QY 661 ArgLysGlyLeuTyrGlnGlyThrGlyArgLeuPhePheGlnPheTyrHisLeuLeuAsn 680  
 Db 2095 AGGAAGGCGCTGTATGAGGATACAGCGCGCTCTTCTGCAATTTTACCACTGCTGAT 2154  
 QY 681 TyrSerArgProLysGlnGlyAspAspArgProPhePheTrpMetPheGlnAsnValAla 700  
 Db 2155 TACTACCGCCCAAGAGAGGATGACCGCGCTTCTTGTGATGTTGAAGATGTTTGA 2214  
 QY 701 AlaMetLysValAlaAspLysArgAspLysSerArgPheLeuGlnCysAsnProValMet 720  
 Db 2215 GCCATGAAGGTGGGCAACAGAGGACATCTCAAGCTTCTGAGAGTGAATCCAGTGTG 2274  
 QY 721 IleAspAlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuPro 740  
 Db 2275 AATTATGCGATCAAGATTTCTGCTGCTCAAGGCGCCGATATCTTGGGGCAACTACCC 2334  
 QY 741 GlyMetAsnArgProValAlaAlaSerLysAsnAspLysLeuGlnAlaAspCysLeu 760  
 Db 2335 GGGATGAACAGGCGCGTATGATGATCAAGATGATTAATCTGAGAGCTGACAGAGCTG 2394

QY 761 GluTyrAsnArgIleAlaLysLeuLysValGlnThrIleThrThrLysSerAsnSer 780  
 Db 2395 GAATACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2454  
 QY 781 IleLysGlnGlyLysAsnGlnLeuPheProValValMetLysGlnGlnGlnGlnGln 800  
 Db 2455 ATCAACAGGGGAAAAACAACCTTCTGCTGTGTCAATGATGCAAGAAAGATGTTTG 2514  
 QY 801 TrpCysThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 820  
 Db 2515 TGGTGCATGAGCTCGAAGAGATTTTGGCTTCTGCTGCTGCTGCTGCTGCTGCTG 2574  
 QY 821 MetGlyArgGlyAlaArgGlnLysLeuGlnGlyArgSerTrpSerValProValIleArg 840  
 Db 2575 ATGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2634  
 QY 841 HisPhePheAlaProLeuLysAspTyrPheAlaCysGln 853  
 Db 2635 CACTCTTCCGCTCGAAGAGCTACTTGTGATGAA 2673  
 RESULT 2  
 AF331857  
 LOCUS  
 DEFINITION Homo sapiens DNA cytosine methyltransferase 3 beta (DNMT3B) mRNA.  
 ACCESSION AF331857  
 VERSION AF331857.1 GI:18033254  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 4335)  
 N.J., Pradhan,S. and Roberts,R.J.  
 Cloning, expression and characterization of human DNMT3 genes  
 JOURNAL  
 TITLE  
 REFERENCE  
 N.J., Pradhan,S. and Roberts,R.J.  
 Direct Submission  
 JOURNAL  
 TITLE  
 Submitted (22-DEC-2000) New England Biolabs, 32 Tozer Road,  
 Beverly, MA 01915, USA  
 FEATURES  
 source  
 1..4335  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..4335  
 /gene="DNMT3B"  
 322..2883  
 /gene="DNMT3B"  
 /codon\_start=1  
 /product="DNA cytosine methyltransferase 3 beta"  
 /protein\_id="A157040.1"  
 /db\_xref="GI:18033255"  
 /translation="MRGDTIRHLNGEEDAGREDLSLVNCAQSDQSDDSPILBAIRTP  
 EIRGRNNSRSLSRKREVSLSLYODLTGDDGDDGDSDPVMPKLFRRRTSBSPA  
 VTRNNNSVSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHS  
 PSSYITLIDTDDTDHGPSSSTPYVALADSDGGGWSGSDGSDGSDGSDGSDGSDGSDG  
 DGERITGLDMKIGTGFSGFPMAMVSKATSRQMSGMRQWEDGSDGSDGSDGSDGSDGSDG  
 LVALGFSQGFNLAIRNKLYSRKAMTHALEKRVAKRTFSPGDSLEDLKPMLE  
 WARGFRPFTIGELKPNNTOPVNVKSRKLESRKLEKTRRRTRRTRRTRRTRRTRRTRR  
 CPARKLNTKNNKNDGDDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDG  
 EGGLCOTGRDRLFLFYDDDGVSCTVCCGRELILCSNTSCGCPCECLELV  
 GTGTAAEAKIOEPMSCMCLPORCHVLRKXDMNVRLQFPISDTGLEAPKLYPA  
 IPAARRRPIRVLSTFGIANGYVLKELGKVKYVASVECSIANGTGKRGNTKY  
 VNDVRNTKNIIEKMGFDLVTCGSPNDLSNVPARKGLYEGTGRLEFEFTHLNYS  
 RPEEGDRPEFMFEENVAMKVGDKSDISRFLSCNCPVMDALIVSAHRRARYFWGNDP  
 GNARPVIASKNDLELDQCLEVNRILAKLKKVQITITKSNISIKGKQDLFPVNVNGKED  
 VLWCTELERIFGSPVHYATDVSNMGRGARQLGRSWSVPVIRLFLPKDYFACE"  
 BASE COUNT 1045 a 1172 c 1167 g 951 t  
 ORIGIN  
 Alignment Scores: 5..29e-234 Length: 4335  
 Pred. No.:

Score: 4566.00 Matches: 853  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-720-086-8 (1-853) x AF331857 (1-4335)

Qy 1 MetLysGlyAspThrArgHisLeuAsnGlyGluGluAspAlaGlyArgGluAspSer 20  
Db 322 ATGAAGGAGACACCAAGGATCTCAATGGAGAGGAGACCCCGGGAGGAGACTCG 381

Qy 21 IleLeuValAsnGlyAlaCysSerAspGlnSerSerAspSerProProIleLeuGluAla 40  
Db 382 ATCTCTGTAACCGGGCCCTGCAGCAGCAGCTCTCCGACTCGCCCCCAATCTCTGAGGCT 441

Qy 41 IleArgThrProGluIleArgGlyArgArgSerSerSerArgLeuSerIysArgGluVal 60  
Db 442 ATCCGCACCCCGAGATCAGAGCCGCAAGATCAAGCTCGCGACTCTCCAAAGAGGAGGTG 501

Qy 61 SerSerLeuLeuSerTyrThrGlnAspLeuThrGlyAspGlyAspGlyGluAspGlyAsp 80  
Db 502 TCCAGTCTGCTAAGCTACACACAGGACTTGCACAGCGATGGCGACGGGGAAGATGGGAT 561

Qy 81 GlySerAspProValMetProLysLeuPheArgGluThrArgThrArgSerGluSer 100  
Db 562 GGCTCTGACACCCCACTCATGCCAAGCTCTTCGGGAAACACAGGACTCTGTTCAAGAGC 621

Qy 101 ProIleValArgThrArgAsnAsnSerValSerSerArgGluArgHisArgProSer 120  
Db 622 CCAGCTCTCGAACTCGAANTAAACAACAGTGTCTCCAGCCGGAGGACACAGGCTTCC 681

Qy 121 ProArgSerThrArgGlyArgGlnGlyArgAsnHisValAspGluSerProValGluPhe 140  
Db 682 CCAGTTCACACCGCGCGGCGGCGCAACCATGTGACAGTCCCGCTGGAGTTC 741

Qy 141 ProIleThrArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProThrProSer 160  
Db 742 CCGCTACCAAGTCCCTGAGACGGCGGCAACAGCATCCGACAGAAACGCGCATGCCGCTCC 801

Qy 161 ProProSerSerTyrLeuThrIleAspLeuThrAspAspThrGluAspThrHisGlyThr 180  
Db 802 CCTCCAGCTCTTACCTTACCATCGACCTCACAGACGACACAGAGGACACATGGGACG 861

Qy 181 ProGlnSerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMet 200  
Db 862 CCCACAGACGACAGTACCCCTACGCCCGCTAGCCCCAGGACAGCCAGCGGGGGCATG 921

Qy 201 GluSerProGlnValGluAlaAspSerGlyAspGlyAspSerSerGluTyrGlnAspGly 220  
Db 922 GAGTCCCGCAGGTGGAGGCACACAGTGGAGATGGAGACATTCAGAGTATCAGGATGG 981

Qy 221 LysGluPheGlyIleGlyAspLeuValTrpGlyIleIysGlyPheSerTrpTrpPro 240  
Db 982 AAGGAGTTTGGAAATAGGGACCTCGTGGGGGAAGATCAAGGGCTTCTCTGGTGGGCC 1041

Qy 241 AlaMetValValSerTrpLysAlaThrSerLysArgGlnAlaMetSerGlyMetArgTrp 260  
Db 1042 GCCATGTGTGTCTTGGAAAGCCACCTCCAAAGCGACAGGCTATGTCTGGCATGCGGPG 1101

Qy 261 ValGlnTrpPheGlyAspGlyLysPheSerGluValSerAlaAspLysLeuValAlaLeu 280  
Db 1102 GTCCAGTGGTGTGGCGATGGCAAGTCTCTCCGAGGTCTCTGCAGACAAACCTGGTGCACTG 1161

Qy 281 GlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLys 300  
Db 1162 GGGGTGTTCAGCCAGCACCTTTAAATTTGGCCACCTTCAATTAAGCTCTCTCTATCGAAA 1221

Qy 301 AlaMetTyrHisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSer 320  
Db 1222 GCCATGTATCCATCTCTGGAGAAAGCTAGGCTGCGAGCTGGCAAGACCTTCCCCAGCAGC 1281

Qy 321 ProGlyAspSerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGlyPhe 340

Db 1282 CCTGGAGACTCATTTGGAGGACACAGCTGAAGCCATGTTTGGAGTGGGCCACAGGGGCTTC 1341

Qy 341 LysProThrGlyIleGlyGlyLeuLysProAsnAsnThrGlnProValValAsnLysSer 360

Db 1342 AAGCCACCTGGATCGAGGGCTTCAAAACCAACACGCAACAGTGGTTAATAAGTCG 1401

Qy 361 LysValArgAlaGlySerArgLysLeuGluSerArgLysTyrGluAsnLysThrArg 380

Db 1402 AAGGTGCGTGGTGCAGGCAGTAGGAATTAGAATCAAGAAATACGAGAAACAAGACTCGA 1461

Qy 381 ArgAthrAlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLeuLys 400

Db 1462 AGACGCAGCTGACGACTCAGCCACCTCTGACTACTGCCCGCCAGCAAGCCCTCAAG 1521

Qy 401 ThrAsnCysTyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluGlnMet 420

Db 1522 ACAAAATGCTATAACAACGCGCAAGACCCGAGGGGATGAAGATCAGAGCCGAGAACAAATG 1581

Qy 421 AlaSerAspValAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerCysGlyVal 440

Db 1582 GCTTCAGATGTTGCCAACAAACAGACAGCCTTGGAAAGATGCTGTTTGTCTTGTGCAGG 1641

Qy 441 LysAsnProValSerPheHisProLeuPheGlyGlyLeuCysGlnThrCysArgAsp 460

Db 1642 AAAAACCCCGCTGCTTCCACCCCTCTCTTTAGGGGGGGCTCTGTCCAGACATGCCGGAT 1701

Qy 461 ArgPheLeuGluLeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThrVal 480

Db 1702 CGCTTCCTTGAAGCTGTTTTACATGATGATGAGATGCTATCAGTCTTACTGCACGTG 1761

Qy 481 CysCysGlyGlyArgGluLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCys 500

Db 1762 TGTCTGAGGGCCGAGAGCTGTGCTTTCAGCAACACAGAGCTGCTGCCGGTGTCTGTGT 1821

Qy 501 ValGluCysLeuGluValLeuValGlyThrAlaAlaGluAlaLysLeuGlnGlu 520

Db 1822 GTGAGTCCCTGGAGGTGCTGGTGGGCACAGACAGCGCGCCAGGCAAGACTTCAGGAG 1881

Qy 521 ProTrpSerCysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgLys 540

Db 1882 CCTCGAGCTGTACATGTGTCTCCCGCAGCGCTGTCTATGGCGCTCTGCGCGCCCGAAG 1941

Qy 541 AspTrpAsnValArgLeuGlnAlaPhePheThrSerAspThrGlyLeuGluTyrGluAla 560

Db 1942 GACTGGAACTGCGCTGCGAGGCTTCTTCCACAGTGACACGGGGCTTGAATATGAAGCC 2001

Qy 561 ProLysLeuTyrProAlaIleProAlaAlaArgArgArgProIleArgValLeuSerLeu 580

Db 2002 CCCAAGCTGTACCTGCCATTTCCCGCAGCCCGAAGCGGCCCATTCAGTCTCTCATTTG 2061

Qy 581 PheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeuGlyIleLysValGlyLys 600

Db 2062 TTTGATGGCATCGCAGCAGGCTACCTAGTCTTCAAGAGATTGGGCAATAAGGTAGGAAAG 2121

Qy 601 TyrValAlaSerGluValCysGluGluSerIleAlaValGlyThrValLysHisGluGly 620

Db 2122 TAGTCTGCTCTGAAGTGTGTGAGGAGTCCATGCTGTGTGAACCGGTGAAGACGAGGGG 2181

Qy 621 AsnIleLysTyrValAsnAspValArgAsnIleThrLysLysAsnIleGluGluTrpGly 640

Db 2182 AATATCAAAATACGTGAACGACGTGAGAAACATCACAAAGAAAAATATTGAAGAAATGGGC 2241

Qy 641 ProPheAspLeuValIleGlyGlySerProCysAsnAspLeuSerAsnValAsnProAla 660

Db 2242 CCATTTGATGTTGGTGTGGCGAAGCCCATGCAACGATCTCTCAAATGTGAATCCAGCC 2301

Qy 661 ArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhePheThrHisLeuLeuAsn 680

Db 2302 AGGAAGGCTGTATAGGGTACAGCGCGGCTCTTCTTCGAATTTTACCACCTGCTGAAT 2361

Qy 681 TyrSerArgProLysGluGlyAspArgProPhePheThrMetPheGluAsnValVal 700



Qy 221 LysGluPheGlyIleGlyAspLeuValTrpGlyLysIleGlyPheSerTrpTrpPro 240  
Db 961 AAGGAGTTTGGAAATAGGGACCTCGTGTGGGAAAGATCAAGGGCTTCTCTGTGTGGGCC 1020  
Qy 241 AlaMetValValSerTrpLysAlaThrSerLysArgGlnAlaMetSerGlyMetArgTrp 260  
Db 1021 GCCATGTTGGTGTCTTGGAAAGCCACCTCCAAAGCGACAGGCTATGTCTGGCATCGGTGG 1080  
Qy 261 ValGlnTrpPheGlyAspGlyPheSerGlyValSerAlaAspLysLeuValAlaLeu 280  
Db 1081 GTCCAGTGGTTTGGCAGTGGCAAGTTCTCCAGAGTCTCTCCAGACAACTGGTGGCACTG 1140  
Qy 281 GlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTrpArgLys 300  
Db 1141 GGGCTGTTTACAGCAGCAGCTTTAAATTTGGCCACCTTCAATTAAGCTCTCTCTATCGAAAA 1200  
Qy 301 AlaMetTrpHisAlaLeuGlnLysAlaArgValArgAlaGlyLysTrpPheProSerSer 320  
Db 1201 GCCATGTACCATGCTCTGGAGAAAGCTAGGGTGGAGCTGGCAAGACCTTCCCCAGCAGC 1260  
Qy 321 ProGlyAspSerLeuGluAspGlnLeuLysProMetLeuGlnTrpAlaHisGlyPhe 340  
Db 1261 CCTGGAGACTCATTTGGAGACCAAGCTGAAGCCCATGTTGGAGTGGGCCCAAGGGGGCTTC 1320  
Qy 341 LysProThrGlyIleGluGlyLysLysProAsnAsnThrGlnProValValAsnLysSer 360  
Db 1321 AAGCCCACTGGGAGCTCAAGGCTCAAAACCAACCAACGCAACCA----- 1365  
Qy 361 LysValArgArgAlaGlySerArgLysLeuGluSerArgLysTrpGluAsnLysThrArg 380  
Db 1366 -----GAGAACCAAGACTCGA 1380  
Qy 381 ArgArgThrAlaAspAspSerAlaThrSerAspTrpCysProAlaProLysArgLeuLys 400  
Db 1381 AGACGACAGCTGACGACTCAGCCACCTCTGACTACTGCCCCCACCACCAAGCGCCTCAAG 1440  
Qy 401 ThrAsnCysTrpAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGlnMet 420  
Db 1441 ACAAAATGTCTATAACACGGCAAGACCGAGGGATGAAGATCAGAGCCGAGAACAAATG 1500  
Qy 421 AlaSerAspValAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerCysGlyArg 440  
Db 1501 GCTTCAGATGTTGCCAACAAACAGAGACCTGGAAGATGGCTGTTTGTCTTGTGGCAGG 1560  
Qy 441 LysAsnProValSerPheHisProLeuPheGluGlyLysLeuGlnThrCysArgAsp 460  
Db 1561 AAAAAACCCGTCCTTCCACCTCTCTTTCAGGGGGGGCTCTGTGACACATGCGCGGAT 1620  
Qy 461 ArgPheLeuGluLeuPheTrpMetTrpAspAspGlyTrpGlnSerTrpCysThrVal 480  
Db 1621 CGCTTCTTGAGCTGTTTACATGATGATGATGATGATGATGATGATGATGATGATGATG 1680  
Qy 481 CysCysGluGlyArgGluLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCys 500  
Db 1681 TCTGCGAGGGCCGAGAGCTGCTGCTTTCAGCAACAGAGCTGCTGCTGCTGCTGCTGCTG 1740  
Qy 501 ValGluCysLeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnGlu 520  
Db 1741 GTGAGTGCCTGGAGTGTGTTGGGCACAGGCACAGCGCCGAGGCCAAGCTTCAGGAG 1800  
Qy 521 ProTrpSerCysTrpMetCysLeuProGlnArgCysHisGlyValLeuArgArgLys 540  
Db 1801 CCCTGGAGCTGTTACATGTTGCTCCCGCAGCGCTGTCTATGGCTCTCTGCGCGCGGAG 1860  
Qy 541 AspTrpAsnValArgLeuGlnAlaPhePheThrSerAspThrGlyLeuGlnTrpGluAla 560  
Db 1861 GACTGGAAAGCTGCGCTGCGAGGCTCTTTCACCAAGTGCACAGGGGCTTGAATATGAAGCC 1920  
Qy 561 ProLysLeuTrpProAlaIleProAlaAlaArgArgArgProIleArgValLeuSerLeu 580  
Db 1921 CCCAAGCTGTACCTGCTCCATCCCGCAGCGCGGCGCCATTCGATTCGCTCATTTG 1980  
Qy 581 PheAspGlyIleAlaThrGlyTrpLeuValLeuLysGluLeuGlyIleLysValGlyLys 600

Db 1981 TTTGTGGCATCGCGACAGGCTACCTAGTCTCTCAAGAGTTGGCATAAAGGTAGGAAG 2040  
Qy 601 TyrValAlaSerGluValCysGluSerIleAlaValGlyThrValLysHisGluGly 620  
Db 2041 TAGTCGCTTCTGAAGTGTGTGAGGAGTCCATTGCTGTGTGAAACCGTGAAGACGAGGG 2100  
Qy 621 AsnIleLysTrpValAsnAspValArgAsnIleThrLysLysAsnIleGluGluTrpGly 640  
Db 2101 AATATCAATACGTGACGACGCTGAGGAACATCAAAAGAAAAATATTTGAAGATGGGC 2160  
Qy 641 ProPheAspLeuValIleGlyGlySerProCysAsnAspLeuSerAsnValAsnProAla 660  
Db 2161 CAAATTGACTTGGTGTATGGGAGCCCATCAACGATCTCTCAATGTGAATCGAGCC 2220  
Qy 661 ArgLysGlyLeuTrpGluGlyThrGlyArgLeuPhePheGluPheTrpHisLeuLeuAsn 680  
Db 2221 AGAAAGGCCCTGTATGAGGGTACAGGCCGGCTCTTCTCGAATTTTACCACCTGCTGAAT 2280  
Qy 681 TyrSerArgProLysGluGlyAspAspArgProPhePheTrpMetPheGluAsnValVal 700  
Db 2281 TACTCACGCCCAAGAGGGGTGATGACCGCGCTTCTTCTGATGTTTGAGAAATGTTGTA 2340  
Qy 701 AlaMetLysValGlyAspLysArgAspIleSerArgPheLeuGluCysAsnProValMet 720  
Db 2341 GCCATGAAGTTGGCAACAGAGGACATCTCACGGTCTCTGGAGTGTAAATCCAGTGATG 2400  
Qy 721 IleAspAlaIleLysValSerAlaAlaHisArgAlaArgTrpPheTrpGlyAsnLeuPro 740  
Db 2401 ATTGATGCATCAAGTTTCTGCTGTACAGGGCCCGATCTCTGGGCGAACCTTACC 2460  
Qy 741 GlyMetAsnArgProValIleAlaSerLysAsnAspLysLeuGluLeuGlnAspCysLeu 760  
Db 2461 GGGATGACAGCGCCGCTGATAGCATCAAGAATGATAAAGCTCGAGCTGCGAGGACTGCTG 2520  
Qy 761 GluTrpAsnArgIleAlaLysLeuLysValGlnThrIleThrLysSerAsnSer 780  
Db 2521 GAATCAATAGTAGAGTACCAAGTTAAAGAAAGTACAGACAATAACCAAGTGCAGACTG 2580  
Qy 781 IleLysGlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysGluAspValLeu 800  
Db 2581 ATCAACAGGGGAAAAACCAACTTTTCCCTGTTGTGATGATGCAATGCAAGAGATGTTTG 2640  
Qy 801 TrpCysThrGluLeuGluArgIlePheGlyPheProValHisTrpThrAspValSerAsn 820  
Db 2641 TGTGCTGACTGAGCTCGAAAGGATCTTTGGCTTTCTGCTGCACTACAGAGCTGTCCAAC 2700  
Qy 821 MetGlyArgGlyAlaArgGlnLysLeuLeuGlyArgSerTrpSerValProValIleArg 840  
Db 2701 ATGGCCGCTGCTGCGCCGAGAGCTGCTGGGAAGTCTCTGGAGGCTGCTGCTCATCCGA 2760  
Qy 841 HisLeuPheAlaProLeuLysAspTrpPheAlaCysGlu 853  
Db 2761 CACCTCTTCGCCCTCTGAAGGACTACTTTCATGTGAA 2799

## RESULT 4

AF156487 3897 bp mRNA linear PRI 05-SEP-1999  
LOCUS Homo sapiens DNA cytosine-5 methyltransferase 3 beta 3 (DNMT3B)  
DEFINITION mRNA, complete cds.  
ACCESSION AF156487  
VERSION AF156487.1 GI:5823165  
KEYWORDS SOURCE  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 3897)  
AUTHORS Xie, S., Wang, Z., Okano, M., Nogami, M., Li, Y., He, W.W., Okumura, K.  
and Li, E.  
TITLE Cloning, expression and chromosome locations of the human DNMT3  
JOURNAL Gene 236 (1), 87-95 (1999)



MEDLINE 99365304  
PUBMED 10433969  
REFERENCE 2 (bases 1 to 3897)  
AUTHORS Xie, S. and Li, E.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUN-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA

FEATURES  
Source Location/Qualifiers

1. 3897  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/map="20q11.2"  
1. 3897  
/gene="DNMT3B"  
115..2427  
/gene="DNMT3B"  
/note="de novo DNA methyltransferase"  
/codon\_start=1  
/product="DNA cytosine-5 methyltransferase 3 beta 3"  
/protein\_id="A053062.1"  
/db\_xref="GI:5823166"  
/translation="MKGDTRLNGEDAGGEDSLVNGACSDSSNDSPILAIRP  
ETRRSSSRISKREVSLSYSTODLGDGDEGDSPTPWPKLFRERTRSESFA  
VTRNNKSVSRERHPSPRSTRGQRNVDSVGFPAIRSLRRATASAGTWPWS  
PSSYLTLDLTDIEDTHGTPQSSSTPYALADDSQGGMESQVEADSGDSSEQ  
DGKEFGIDLVGKIKGFSWMPAMVSMKATSKROMGKTPQWFGDGFSEVADK  
LVALGFESQHNFLATFNKLVYRKAMTHALEKARVRAKTPPSPGSLDLKPMLE  
MAHGFEPTGIEGLKPNNTQPEKTRRTADATSPYCPAPRIKTNKCNKRDGD  
EDDSRROMADVANNKSSLEDGLSCGRKAPVSPHPIFEGLCOTCDRLLELYKMD  
DDGISTCTVCCGRELDCSNTSCCFCEVECLEVLVGTAAEALQSPWSCYML  
PQRGVLRRKRMVRLQAFPTSDGLEAEALYPAIPARRRPIRVLSTFDGAT  
GYLVLELGIKVKYVASEVCEESIAGVTYHNGNIYVVDNITTKNLENGPDL  
VIGSEPCNDLSNVPARKGLYEGTGRLFFEFYHLNLRKREGDPRFFPMFENVAM  
KVGDKDISRFLRCPNPMIDAIKYSAAHRAIRYFMGNIPGNRIFFPVHTTDSNMGR  
GAQKILGRMSVPVIRHLPALPKDYACE"

BASE COUNT 951 a 1039 c 1033 g 874 t  
ORIGIN

Alignment Scores:  
Pred. No.: 2,17e-208 Length: 3897  
Score: 4083.50 Matches: 770  
Percent Similarity: 90.27% Conservative: 0  
Best Local Similarity: 90.27% Mismatches: 0  
Query Match: 89.43% Indels: 83  
DB: Gaps: 2  
US-09-720-086-8 (1-853) x AF156487 (1-3897)

QY 1 MetLyseGIYAaPTThrArgHIEleuAaNGIyGIuGIuAaPAlaGIyGIYAArgGIuAaSPSer 20  
Db 115 ATGAAGGAGACACACAGGACATCTCAATGAGAGAGACGCGCGGAGGAGGAAAGACCTCG 174  
QY 21 ILeuValAaNGIyAaCySeSaSPGInSeSaSPSerProProlIEleuGIuAa 40  
Db 175 ATCTCTGTCACAGCGGCTCGACAGCACAGTCTCCGACTGCGCCCAATCTCGAAGGCT 234  
QY 41 ILeaRgThrProGIuIleArgGIYAArgSerSerSerAaGleuSerIyAArgGIuVal 60  
Db 235 ATCCGACACCCCGAGATCAAGAGCCGGAAGATCAAGCTCGCACTCTCCAAAGGAGAGGTG 294  
QY 61 SerSerIeuIeuSerIyThrGIuAaSPleuThrGIYAaSPGIyGIuAaSPGIYAaSP 80  
Db 295 TCCAGTCTGCTAAGCTACACACAGACCTTGACAGCCATGCGACGAGGAGGAAATGGGAT 354  
QY 81 GYSeSaSPThrProValMetProIySeuIeuAaArgGIuThrArgThrArgSerGIuSer 100  
Db 355 GGCTCTACACCCAGATCATCCAAAGCTCTTCGCGGAAACACAGACTCGTTCAAGAAAGC 414  
QY 101 ProIaValaArgThrArgAaAaAaAaSerIySerSerAaArgGIuAaGHIaArgProSer 120  
Db 415 CCAGCTGTCCAACTCCGAATTAACAACAGTGTCTCCAGCCGAGGAGGACACAGGCTTCC 474

QY 121 ProARgSerThrArgGIYAaRGInGIYAArgAaNHISValAaSPGIuSerProValGIuPhe 140  
Db 475 CCAGTTCACACCCAGAGCCGACAGGCGCCGACCATGTGACGAGTCCCCCGTGAAGTTC 534  
QY 141 ProIaThrArgSerIeuAaArgAaArgAlaThrIaSerIaGIyThrProTrpProSer 160  
Db 535 CCGGCTACCAAGTCTCCGAGCGGCGGACACACATCGGAGGAAAGCCATGGCCGCTCC 594  
QY 161 ProProSerSerIyIleuThrIleAaSPleuThrAaSPThrGIuAaPThrHIEGIyThr 180  
Db 595 CCTCCAGCTCTTACCTTACATGACCTCAGACAGCACAGAGACACACATGGGAGC 654  
QY 181 ProGIuSerSerSerThrProTyraIaArgIeuaIaGIuAaSPSerGIuNGIyGIuMet 200  
Db 655 CCCCAGAGACAGACATACCCCTTACGCCGCTTACCCAGACAGCACAGAGGAGGCGATG 714  
QY 201 GIuSerProGIuValaGIuAaSPSerGIYAaSPGIyAaSPSerSerGIuTyrgIuAaSPGIy 220  
Db 715 GAGTCCCGCAGGTGGAGGACGACAGTGAAGTGAACAGTTGAGATACAGATATGGG 774  
QY 221 LyseGIuPheGIyIleGIYAaSPleuValITripGIyLySIIeLySGIyPheSerITripPro 240  
Db 775 AAGAGATTGGAATRAGGAGACTGTGTGGGAAAGATCAAGGCTTCTCTGTGAGGCC 834  
QY 241 AlaMetValaIaSerITripLyAaIaThrSerIyAaArgIuAaMetSerGIyMetAArgITrip 260  
Db 835 GCGATGGGTGGTCTTGGAAGGCCACTCCAAAGGACAGGCTATGTCTGGCATGCGGTGG 894  
QY 261 ValGIuThrPheGIYAaSPGIyLyAaPheSerGIuValaSerIaAaSPlyIleuValaIleu 280  
Db 895 GTCCAGTGTGTGGCGATGGCAGATTTCCGAGGCTCTCGACAAACGTGTGGCACTG 954  
QY 281 GIyIeuPheSerGIuNHISpHeaNHIEuAlaThrPheAaNHISyIleuValaIaSerITyAArgLyS 300  
Db 955 GGGTGTTCACCCGACCTTAAATTTGGCCACTTCAATAGCTCGTCTCTTATCGAANA 1014  
QY 301 AlaMetTyriIaIeGIuGIuLyAaIaArgValaArgIaGIyLyThrPheProSerSer 320  
Db 1015 GCCATGTACCATGCTCTGAGAAAGCTAGGGTGGAGCTGGCAAGACTTCCACGAGCAGC 1074  
QY 321 ProGIyAaSPSerIeuGIuAaSPGIuIleuLyProMetIeuGIuITripAlaHIEGIyPhe 340  
Db 1075 CCTGAGACATCATTTGAGGAGACACAGCTAAGCCCTGTTGGATGGGCGCACAGGGGGCTTC 1134  
QY 341 LyAaProThrGIyIleGIuGIuLyIeuLyProAaAaNHInIaProValaIaNHISer 360  
Db 1135 AAGCCACCTGGAGTGAAGGGCTCTAAACCCAAACACGCAACCA----- 1179  
QY 361 LySeValaArgAArgIaGIySerAArgLyIeGIuSerAArgLyTyrgIuAaNHISerThrArg 380  
Db 1180 -----GAGAACAAAGACTCGA 1194  
QY 381 ArgAArgThrAlaAaPaSPSerAlaThrSerAaSPTyrcysProAlaPolysAArgIeuLyS 400  
Db 1195 AGAGCGACAGCTGACGACTAGCACCTCTGACTACAGCCCGCACCAAGCGCCTCAAG 1254  
QY 401 ThrAaNHISyTyriAaAaNGIyLyAaSPaArgGIYAaSPGIuAaSPGIuSerAArgGIuIleMet 420  
Db 1255 ACAAATTGCTATTAACAACGGAAGAACCGAGGGAGTAAAGTCAAGACCGGAAACAATG 1314  
QY 421 AlaSerAaSPValaIaAaAaNHISerSerIeuGIuAaSPGIyCysIeuSerCyGIyAArg 440  
Db 1315 GCTTCAGATGTGTCACAAACAAAGACAGCAGCTGGAAGATGGCTTTGTCTGTGGCAGG 1374  
QY 441 LyAaAaProValaIaSerPheHISProIeuPheGIuGIyIeuCyGIuThrCyAArgAaSP 460  
Db 1375 AAAAACCCTGTCCTTCCACCCCTCTTTGAGGGGGGGCTGTGTGACAGACTGCGGGAT 1434  
QY 461 ArgPheIeuGIuIeuPheTyriMetTyriAaSPaSPaSPGIyTyrgIuNHISerTyrcysThrVal 480  
Db 1435 CGTTCCTTGAAGCTGTTTACATGATGATGACATGAGTGTCTTACGTGACATGTCG 1494  
QY 481 CyAeGIuGIuGIyAArgGIuIleuIeuCySeSaNHISerCySaArgCyAePheCyS 500

Db	2386	CACCTCTTCGCCCTCTGAAGGACTACTTTCATGTGAA	2424
RESULT 5			
AF068626			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
AUTHORS			
TITLE			
JOURNAL			
REMARK			
COMMENT			
FEATURES			
source			
gene			
misc_feature			
CDS			
1495		TGCTGGAGGCGGAGAGCTCTGCTTTGAGCAACACAGAGTGTGCGGTGTTCTGT	1554
501	ValGluCysLeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnGlu		520
1555	GTGGAGTGTCTGGAGTGTCTGTTGGGACAGGACAGCGCGCGGCGCAAGCTTTCAGGAG		1614
521	ProTrpSerCysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgLys		540
1615	CCCTGGAGCTGTACATGTGTCTCCCGCAGCGTGTATGCGGTCTCTGCGGCGCGAAG		1674
541	AspTrpAsnValArgLeuGlnAlaPhePheThrSerAspThrGlyLeuGluTyrGluAla		560
1675	GACTGGAAAGTGGCGCTCGACGCGCTCTTCCACAGTGACACGCGGCTTGAATACGAAGCC		1734
561	ProLysLeuTyrProAlaIleArgAlaArgArgProIleArgValLeuSerLeu		580
1735	CCCAAGCTGATCCCTGCCATTCCTCCGAGCGCCGAGCGGCCCATTCGAGTCTCTCATTTG		1794
581	PheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeuGlyIleLysValGlyLys		600
1795	TTTGATGGCATCGCACAGGCTACTAGTCTCAAAGAGTTGGGCATAAAGTAAAGGAAAG		1854
601	TyrValAlaSerGluValCysGluGluSerIleAlaValGlyThrValLysHisGluGly		620
1855	TAGTCGCTTCTGAAAGTGTGAGAGTCCATTTGCTGTTGGAAACCGTGAACAGCAGGGG		1914
621	AsnIleLysTyrValAsnAspValArgAsnIleThrLysLysAsnIleGluGluTrpGly		640
1915	AATATCAATACGTGAACGAGTGAAGAACATCACAAAGAAATATTTGAAGAATGGGCG		1974
641	ProPheAspLeuValIleGlyLysProCysAsnAspLysSerAsnValAsnProAla		660
1975	CCATTTGACTTGGTGAATGGCGGAAGCCCATGCAACGATCTCTCAATGTGAATCCAGCC		2034
661	ArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPheTyrHisLeuLeuAsn		680
2035	AGGAAAGGCTGTATGAGGTACAGGCGGCTCTTCTCGAATTTTACCACCTCTCGAAT		2094
681	TyrSerArgProLysGluValAspArgProPhePheTrpMetPheGluAsnValVal		700
2095	TACTCAGCCCCNAGAGGGTGTATGACCGGCGCTCTTCTGAGTTTGAAGATGTGTA		2154
701	AlaMetLysValGlyAspLysArgAspIleSerArgPheLeuGluCysAsnProValMet		720
2155	GCCATGAGTTGGCGACAGAGGATCTCACGGTTCTCGAGTGTATCCAGTGTG		2214
721	IleAspAlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuPro		740
2215	ATTGATGCCATCAAAAGTTTCTGCTGCTCACAGGCGCGATCTTCTGGGGCAACCTACCC		2274
741	GlyMetAsnArgProValIleAlaSerLysAsnAspLysLeuGluLeuGlnAspCysLeu		760
2275	GGGATGAAC-----		2283
761	GluTyrAsnArgIleAlaLysLeuLysLysValGlnThrIleThrLysSerAsnSer		780
2283	-----		2283
781	IleLysGlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysGluAspValLeu		800
2283	-----		2283
801	TrpCysThrGluLeuGluArgIlePheGlyPheProValHisTyrThrAspValSerAsn		820
2284	-----AGCATCTTTGGCTTTCTCTGTGTCACACTACACAGAGCTGTCCAAC		2325
821	MetGlyArgGlyAlaArgGlnLysLeuLeuGlyArgSerTrpSerValProValIleArg		840
2326	ATGGGCGGTGGTGGCGGACAGAGCTCTCTGGGAGGCTCTCTGAGGCTGCTGTATCCGA		2385
841	HisLeuPheAlaProLeuLysAspTyrPheAlaCysGlu		853

BASE COUNT	1061	a	1125	g	950	t

BASE COUNT 1061 a 1059 c 1125 g 950 t

Db	2386	CACCTCTTCGCCCTCTGAAGGACTACTTTCATGTGAA	2424
RESULT 5			
AF068626			
LOCUS			
DEFINITION			
AF068626			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REMARK			
COMMENT			
FEATURES			
source			
gene			
misc_feature			
CDS			

AF068626	4195 bp	mRNA	linear	ROD 06-DEC-1999
Mus musculus DNA cytosine-5 methyltransferase 3b1 (Dmmt3b) mRNA,				
alternatively spliced, complete cds.				
AF068626				
AF068626.2	GI:6449469			
Mus musculus				
Mus musculus				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
1 (bases 1 to 4195)				
Okano, M., Xie, S. and Li, E.				
Cloning and characterization of a family of novel mammalian DNA				
(cytosine-5) methyltransferases				
Nat. Genet. 19 (3), 219-220 (1998)				
98324766				
9662389				
2 (bases 1 to 4195)				
Xie, S., Okano, M. and Li, E.				
Direct Submission				
Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,				
Charlestown, MA 02129, USA				
3 (bases 1 to 4195)				
Okano, M., Chijiwa, T., Sasaki, H. and Li, E.				
Direct Submission				
Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,				
Charlestown, MA 02129, USA				
Sequence update by submitter				
On Nov 18, 1999 this sequence version replaced gi:3327979.				
Location/Qualifiers				
1..4195				
/organism="Mus musculus"				
/db_xref="taxon:10090"				
/chromosome="2"				
/map="84.0 cM"				
1..4195				
/gene="Dmmt3b"				
1..4195				
/gene="Dmmt3b"				
/notes="similar to EST sequences deposited in GenBank"				
Accession Numbers AA11694, AA11999, AA17277, AA210568				
AA07106, and AA575617"				
269..2848				
/gene="Dmmt3b"				
/function="de novo DNA methylation"				
/note="alternatively spliced product; contains Cys-rich				
region; C-terminal region is similar to corresponding				
region of DNA cytosine-5 methyltransferase 3a"				
/codon_start=1				
/product="DNA cytosine-5 methyltransferase 3b1"				
/protein_id="AAC40178.2"				
/db_xref="GI:6449470"				
/translation="MKGDSRLNEEAGSYBECIIVNGNFSQSSDTKDAPSPVLE				
ALCTPVCPTETRRSSRLSKREVSSLLNYTQMTGDRDDDDGDSGLMPK				
LYRETKDTRRSSEPAVTRHNSGTSLSERASPRITRGQRHHVVOEYVPEPATR				
SRRRASSTSPWSPASVDPMEEVTPKSVSTPSVLSQDQEGMDTQVDAESRD				
GSTYQDDKEFGIGDLVWGKIGKFSWPMVVKATSKROAMPGRHWVQDGGKF				
SETSDAKLVALGLFSOHFNATFNKLVSRYKAMVHTLEKARVRACKTFSSPSGSLD				
OLKPMLEWAGHKFTGTIEGLKNKQPPVVKSKVRSRNLPRRRKSRRTTN				
DSAAESPPKRLKTSYGGKDRGEDEBERMASEVTNNKLNEDRLCLSCGKNPV				
PHLEPGGLCSQDRDFLEFLFYMDGQVSYCTVCEGREILLCSNCCRCFCVSC				
LEVLVAGTASDAKLQRPWSCYMLQRCRGVLRKKNMRLQDFTTDFLEFEP				
PKLYPAIAKRRPIRLVLDGATGTYLVKLKELGIKVEKYIASVCAESTAVGTVKH				
EGQIKYVNDVRKIKNLEEGPDDLVIGSCNDLSNNPARKGLVGGTGRLEFEY				
HLNLTYPKEGDNRPFPMFENVVMKVDKDISREFLACHPMIDAIKYSAHARY				
FNGNLPGMNRPMASKDKLELODCLFEFSTAKLKVQTIITKNSIRQGNOLPPV				
MNGKDDVLWCTELERIFGFPAAHYTDVSNMCRGNRQKLGRSWSVPIHLPFLKDYF				
ACE"				

ORIGIN

### Alignment Scores:

4.33e-191

3759.00

88.63%

82.32%

82.33%

10

1  
2  
3  
4  
5  
6  
7

X AF06862

**Intermittent**

# hTARGHISLE

11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526  
 527  
 528  
 529  
 530  
 531  
 532  
 533

GCAGACATCT

1vA]aCvase

US-09-720-086-8 (1-853) X AF068626 (1-4195)

Qy	1	MeIyvgIlyAspThrIaArghIstleuAangIylngIuIAspAlaGlyIyArghIuAspSer	20
Db	269	ATGAAGGAGAGCAGCAACACTTGAATGAAGAAGAGGGTCCAGCGGGTATGAGAGAGTGC	328
Qy	21	IleuValAsngIyAlaCySerAspGlnSerSerAsp-----SerPro	35
Db	329	ATTATCGTATATGGGAACCTTCAGTGCACAGTCTCTCAACAACGAAGATGCTCCCTCACCC	388
Qy	36	ProIleuGluAlaIle-----ArgThrProGluIleArgIyIyArg	50
Db	389	CCAGCTCTGGAGGCAACTCTGCACAGAGCCAGTCTGCACACCAAGACCGAGGCGGACAG	448
Qy	51	SerSerSerArgIleuSerIlyIyArghIuValSerSerIleuIleuSerIlyThrIAlAspIleu	70
Db	449	TCAAGCTCCCGCGCTGTCTAAGAGGAGGTCTCCAGCGCTTCTAATTAACGACGACGACTG	508
Qy	71	ThrIyAspGlyAspGly-----GluAspGlyAspGlySerAspThrProAla	86
Db	509	ACAGAGATGGACACACAGACTGATGAAGTAAATGAATGGAAATGGCTTGATTT---CTA	565
Qy	87	MetProIyIleuPheArgIu-----ThrArgThrArgSerGlnSerProAlaVal	103
Db	566	ATGCCAAGCTCACCCGTGAGCAACAGACACCCAGGACCGCTCTGAAAGCCCGGCTTCC	625
Qy	104	ArgThrIyArgIleuAsnAsnSerValSerSerArgIuIyArgIleuArgProIySerProArgSer	123
Db	626	CGAACCCGACATAGCAATGGGACCTCCAGCTTGGAGGCAAGAGCCCTCCCCCAAGATC	685
Qy	124	ThrArgIyIyArgIleuIyArgAsnHisValAspGlnSerProValGluPheProAlaThr	143
Db	686	ACCCGAGCTCGGCGAGGCGCCGACACATGTGCAGAGATACCCTGTGAGATTCTCCGGCTACC	745
Qy	144	ArgSerIleuArgIyArgIleuAlaThrAlaSerAlaGlyIyThrProIyProIySerProIySer	163
Db	746	AGGCTCGGAGAGCTCGAGCATGCTCTTCACGACAGACCCATGTCATCCCTCCACAGC	805
Qy	164	SerIyIleuThrIleAspIleuThrAspAspThrIyAspThrHisGlyIyThrProGlnSer	183
Db	806	-----GTCGACTTCATGAGAAAGATG-----ACACCTTAAGAGC	838
Qy	184	SerSerThrProIyAlaIyGluAlaGlnAspSerGlnGlnGlyIyMetGlnSerPro	203
Db	839	GTCAGTACCCCATCAGTTGACTTGAGCCAGAGATGAGATCAGAGCGGTATGGATGCACCA	898
Qy	204	GlnAlaGlnAlaAspSerGlyIyAspGlyAspSerSerGlyIyGlnAspGlyIyGlnIleuPhe	223
Db	899	CAGGTGATGCGAGACACAGATGAGACGACGACACAGATATCAGATATTAAGAATTT	958
Qy	224	GlyIleGlyAspIleuValIyProGlyIyIyIyGlyIyPheSerIyIyProAlaMetVal	243
Db	959	GGATATAGGTGACCTCGTGGGGGAAGATCAAGGGCTTCTCCGTGGGCTGCTCATGGTG	1018
Qy	244	ValSerIyIyAlaIyThrSerIyIyArgGlnAlaMetSerGlyIyMetArgIyIyAlaGlnIyPhe	263
Db	1019	GTGTCTCGAAAGCCACTCCAGCGCAGAGCCATCCCGGAATGGCGTGGGTACAGTGG	1078
Qy	264	PheGlyIyAspGlyIyIyPheSerGlyIyValSerAlaAspIyIleuValAlaIleuGlyIyPhe	283
Db	1079	TTTGGTATGCGCAGATTTTCTGAGATCTTCGCTGACAAACTGGTGGCTTGGGCGCTGTTCC	1138
Qy	284	SerGlnHisPheIleuValaIyThrPheAsnIyIleuValSerIyIyArgIyAlaMetIyTr	303

Db 1139 AGCCAGCACTTTAACTGCGCTACCTTCATTAAGCTGGTTCTTTAAAGAAAGCCATTGAC 1198

QY 304 H1SAlaLeuGluValAspAlaArgValArgAlaGlyLysThrPheProSerSerProGlyLysp 323

Db 1139 CACACTCTGGAGAAAGCCAGGGTTCCAGCTGGCAGACCTTCTCCAGAGCTCTGGAGAG 1258

QY 324 SerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGlyPheLysProThr 343

Db 1259 TCACCTGGAGAGCCAGCTGAAGCCCATCTGAGATGGGGCCACAGGTGGCTTCAAGCTACT 1318

QY 344 GlyIleGluGlyLeuLysProAsnAsnThrGlnProValValAsnLysSerLysValArg 363

Db 1319 GGGATCCAGGGCCCTCAAAACCAACCAAGCAACAGTGGTTAAATTAATCTCAAGGTGGT 1378

QY 364 ArgAlaGlySerArgLysLeuGlnSerArgLysTyrGluAsnLysThrArgArgThrThr 383

Db 1379 CGTTTCACACACTAGAACTTGAACCCAGAGAGCCGGAGAACAAAGTCCAGAACGCACA 1438

QY 384 AlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLeuLysThrAsnCys 403

Db 1439 ACCAATGACTCTGGCTGCTTGAG--TCCCCCCAACCAAGCCCTCAAGCAAAATAGC 1495

QY 404 TyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluGlnMetAlaSerAsp 423

Db 1496 TAT--GCGCGGAAGGACCCAGGGAGGATGAGAGAGCCGGAAGCGATGGCTTCTGAA 1552

QY 424 ValAlaAsnAsnLysSerSerLeuGluAspGlyCysLysSerCysGlyArgLysAspPro 443

Db 1553 GTCAACCAACAAGGGCAATCTGGAAAGACCGCTGTTTCTCTGGAAAGAGAAACCTT 1612

QY 444 ValSerPheHisProLeuPheGluGlyGlyLeuCysGlnThrCysArgAspArgPheLeu 463

Db 1613 GTGTCCTTCCACCCCTCTTGAGGGTGGGCTCTGACAGTTGCGGGATTCGCTTCTTA 1672

QY 464 GluLeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThrValCysCysGlu 483

Db 1673 GAGCTCTTCAACATGATGATGAGAGAGCGCTATCACTCTCAAGCCGTGGCTGTGAG 1732

QY 484 GlyTyrGluLeuLeuLeuCysSerAsnThrSerCysArgCysPheCysValGluCys 503

Db 1733 GGCGGTGAACGTGCTGTCAGATGACACACAGTGTCTGAGATGCTTCTGTGTGAGTGT 1792

QY 504 LeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnGluProTrpSer 523

Db 1793 CTGGAGAGTGGTGGGGCGAGGCCAGCTCAGAGAGGACCACTCCAGAAACCTCGAGAC 1852

QY 524 CysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgLysAspTrpAsn 543

Db 1853 TGTATATATGTCCTCTCCCTCAGCGCTCCATGGGGTCTCTCCAGCAGGAAAGATTGGAAC 1912

QY 544 ValArgLeuGlnAlaPhePheThrSerAspThrGlyLeu--GluTyrGluAlaProLys 562

Db 1913 ATGGCGCTGCAAGACTTCTTCACTACTGTATCTGACTGGAGAAATTTGAGCCACCCAAAG 1972

QY 563 LeuTyrProAlaIleProAlaAlaArgArgArgProIleArgValLeuSerLeuPheAsp 582

Db 1973 TTGTACCCAGCAATCTCTGAGGCCAAAGAGGCCATTAGAGTCTCTCTCTTTGAT 2032

QY 583 GlyIleAlaThrGlyTyrLeuValLeuLysGluLeuGlyIleLysValGlyLysTyrVal 602

Db 2033 GGAATTGGACAGGGGTACTTGGTGTCTCAAGAGATTGGGTATTAAGTGGAAAAATGACATT 2092

QY 603 AlaSerGluValCysGluGlnSerIleAlaValAlaGlyThrValLysHisGluLysAsnIle 622

Db 2093 GCGTCCGAAGCTGTGAGAGTCCATTCGCTGTGGGAACGTTTAAACATGAAGAGGCCAGATC 2152

QY 623 LysTyrValAsnAspValArgAsnIleThrLysLysAsnIleGluGluTrpGlyProPhe 642

Db 2153 AAATATGTGCATAGCGTCCGGAATATCCACCAAGAAAAAATTTGAAGATGGGGCCCGTTC 2212

QY 643 AspLeuValIleGlyLysProCysAsnAspLeuSerAsnValAsnProAlaArgLys 662

```
Db 2213 GACTTGGTGGTGGAGCCCATGCAATGATCTCTTAACGTCAATCTGTCGCGCAAA 2272
Qy 663 GlyLeuTyGluGlyThrClyArgLeuPheGluPheTyHisLeuLeuAsnTySer 682
Db 2273 GGTATATATGAGGACACAGGAAGCTCTTCTCGAGTTTACCACCTTCTGTAATATACC 2332
Qy 683 ArgProLyGluGlyAspAspArgProPhePheTrpMetPheGluAsnValValaMet 702
Db 2333 CGCCCAAGAGGGGCAACACGTCATCTTCTGATGTTTGAAGATTTGTGGCCATG 2392
Qy 703 LysValGlyAspLyArgAspLySerArgPheLeuGluCysAsnProValMetIleAsp 722
Db 2393 AAAGTGAATGACAAAGAAAGACATCTCAAGATTTCTGGCATGTAAACCCAGTATGATCGAT 2452
Qy 723 AlaileYValSerAlaAlaHisArgAlaArgTyPheTrpGlyAsnLeuProGlyMet 742
Db 2453 GCCATCAAGTGTCTGCTGCTCACAGGGCCCGGTACTCTTGGGGGTAAACCTTACCCGGAATG 2512
Qy 743 AsnArgProValIleAlaSerLyAsnAspLyLeuGluLeuGlnAspCysLeuGluTy 762
Db 2513 AACAGGCCCGTGTATGCTTCAAGAAATGATAAGCTCGAGCTGCAGGACTGCTCGGATTC 2572
Qy 763 AsnArgIleAlaLeuLeuLeuValGlnThrIleThrThyLysSerAsnSerIleLys 782
Db 2573 AGTAGCAGCAAGAAAGTAAAGAAAGTGCACACAAATAACCAACCAAGTCCATCAGA 2632
Qy 783 GlnGlyLeuAsnGlnLeuPheProValValMetAsnGlyValGluAspValIleTrpCys 802
Db 2633 CAGGGCAAAAACACGCTTTCCCTGTAGTCATGANTGGCAAGGACGAGCTTTGTGGTGC 2692
Qy 803 ThrGluLeuGluArgIlePheGlyPheProValHisTyThrAspValSerAsnMetGly 822
Db 2693 ACTGAGCTCGAAGAAATCTTGGCTTCCCTGCTCACTACACGAGTGCTTCAACATGGGC 2752
Qy 823 ArgGlyAlaArgGlnLysLeuLeuGlyArgSerTrpSerValProValIleArgHisLeu 842
Db 2753 CGCGGGCCCGCTCAGAGCTGCTGGCGAGGCTCTCGAGGTACCGGTATCAGACACCTG 2812
Qy 843 PheAlaProLeuLyAspTyPheAlaCysGlu 853
Db 2813 TTTGGCCCTTGAAGGACTACTTTGCTGTGAA 2845

RESULT 6
LOCUS AF151973 4223 bp mRNA linear ROD 08-JUN-2000
DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 5
ACCESSION AF151973
VERSION AF151973.1 GI:8347127
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4223)
AUTHORS Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
Shen,Y.
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing
JOURNAL Isoforms in mouse embryonic tissue
REFERENCE 2 (bases 1 to 4223)
AUTHORS Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
Shen,Y.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular
Biology, Institution of Basic Medical Sciences, Chinese Academy of
Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,
P.R.China
FEATURES
Location/Qualifiers
1..4223
/organism="Mus musculus"
/strain="KM"
/db_xref="taxon:10090"
```

```
/dev_stage="8-9 day old embryo"
1..4223
/genes="Dnmt3b"
297..2876
/genes="Dnmt3b"
/function="methylates cytosine in DNA"
/notes="alternatively spliced"
/codon_start=1
/products="DNA cytosine-specific methyltransferase isoform 5"
/protein_id="AAF74519.1"
/db_xref="GI:8347128"
/translation="MKGDSRHLNBEAGSYECIIIVGNFSDQSDTKDAPSPVLE
AICTEPTCTPSTRRRSSRLSKREVSLLNYTQDMTGDGDDDDVDDGNSDILMPK
LITRETKDTRTSESAPVATRHNSNGTSLERQASPRITRGRQRHHVQVEYEVFPAIR
SRRRASSASTPMSPASVDVFMEEVTPKSVPMVSVKATSKRAMFGMRWVQMGDGI
GDSYEDDDKEFGIDLVGKIKGFSWMPAMVSVKATSKRAMFGMRWVQMGDGI
SEISDKLVAGLFSQHFNLTNKLKVPVYKAMHYHLEKARVRAGKTFTSSPGSLSL
QLKPMLEWAHGGFKFTIEGLKPKKKQPVVVKSKVRRSDRNLEPRRREKSRRTTN
DSAAESPPLKRLKTNISYGGKRGDEDEESRERMASEVTNNGNLLCSNCCRCFVFC
FHLFEGGLQSCDRDRFLELFYMYDEGQSYCTVCCGRELKLLCSNCCRCFVFC
LELVGAGTAEDAKLQEPWCYMLPQCHGVLRRLKRMRLQDFFTTDPDLBEFEP
PKLYPAIPAAKARPRIIVLSLPDGIATGILVKELGIVKVEYIASEVCAEIAVGTVKH
EGQIKYVNDVRKI TKKNIIEGWPFDLVIGGSPCNDLSNVNPAKGLYEGTGRLPPEFY
HLNLYTRPKGDNRPFFMFPENVMKNDKDISRLACNPVMDALIKVSAARARY
PWGNLPGMNRPVMAKNDKLELODCLEFSRTAKLKKVOTITTKSNSIQGNOLFPVV
MNGKDDVLWCTELERIFGFPAPHYTDVSNMGRGAROKLGRSWSVPVIRHLPALPKDYF
ACE"
```

BASE COUNT 1051 a 1084 c 1147 g 941 t

ORIGIN

Alignment Scores:

Pred. No.:	1..68e-190	Length:	4223
Score:	3748.00	Matches:	716
Percent Similarity:	86.52%	Conservative:	55
Best Local Similarity:	82.20%	Mismatches:	70
Query Match:	82.08%	Indels:	30
DB:	10	Gaps:	10

US-09-720-086-8 (1-853) x AF151973 (1-4223)

Qy 1 MetLysGlyAspThrArgHisLeuAenGlyGluAenGlyArgGlyArgGluAspSer 20

Db 297 ATGAAGGAGACACACACATCTGAATGAAGAAGAGGTGCCAGCGGTATGAGGATGC 356

Qy 21 IleLeuValAenGlyAlaCysSerAspGlnSerSerAsp-----SerPro 35

Db 357 ATTATCGTTAATGGAACTTTCAGTGACCATCTCTCAGACACGAAGGATGCTCCCTCACCC 416

Qy 36 ProfileLeuGluAlaIle-----ArgThrProGluIleArgGlyArgGlu 50

Db 417 CCAGTCTTTGGAGGCAATCTGCACAGAGCCAGTCTGCACACAGAGACGAGCCGACG 476

Qy 51 SerSerSerArgLeuSerLyArgGluValSerSerLeuLeuSerTyThrGlnAspLeu 70

Db 477 TCAAGCTCCCGGCTGCTTAAGAGGAGGTCTCCAGCTTCTGAATTACACGAGACATG 536

Qy 71 ThrGlyAspGlyAspGly-----GluAspGlyAspGlySerAspThrProVal 86

Db 537 ACAGAGATGGACAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 593

Qy 87 MetProLysLeuPheArgGlu-----ThrArgThrArgSerGluSerProAlaVal 103

Db 594 ATGCCAAAGCTACCCCTGGAGACCAAGGACACGAGACGCGCTCTGAAGCCCGGCTGTC 653

Qy 104 ArgThrArgAenAenAenSerValSerArgGluArgHisArgProSerProArgSer 123

Db 654 CGAACCCACATAGCAATGGACCTCCAGCTTGGAGAGGCAAGAGAGCTCCCCCAATC 713

Qy 124 ThrArgGlyArgGlnGlyArgAenHisValAspGluSerProValGluPheProAlaThr 143

Db 714 ACCCGAGGTGCGAGGCGCCACCATCTGTGCAGGAGTACCTCTGTGAGTGTTCGGCTACC 773

QY 144 ArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProTyrProSerProProSer 163  
 Db 774 AGGTCGCGAAGCGTCGACATCGCTTCAGCAAGCAGCCATGCTATCCCTGCCAGC 833  
 QY 164 SerTyrLeuThrIleAspLeuThrAspAspThrGluAspThrHisGlyYThrProGlnSer 183  
 Db 834 -----GTCGACTTCATGAGAAAGTGC-----ACACCTTAAGGC 866  
 QY 184 SerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMetGlnSerPro 203  
 Db 867 GTCAGTACCCCATGACTGATCTGAGCCAGATGATGATGAGAGGTATGATACACCA 926  
 QY 204 GlnAlaGlnAlaAspSerGlyAspGlyAspSerSerGlyTyrGlnAspGlyLysGlnPhe 223  
 Db 927 CAGGTGATGAGAGAGATATATGAGACAGCAGAGTATCAGAGATGAAGAAAGATT 986  
 QY 224 GlyIleGlyAspLeuValTyrGlyLysIleLysGlyPheSerTyrProAlaMetVal 243  
 Db 987 GGAATTGGTGACCTCGTGTGGGAAAGATCAAGGCTTCTCTGCTGCTGCATGGGTG 1046  
 QY 244 ValSerTyrPsalIaThrSerLysArgGlnAlaMetSerGlyMetArgTyrValGlnTyr 263  
 Db 1047 GTGCTCTGGAAGCACCCTCCAGGCAAGGCCATGCCGGAAATGCCGTGGTACAGTGG 1106  
 QY 264 PheGlyAspGlyLysPheSerGlyLysSerAlaAspLysLeuValAlaLeuGlyLeuPhe 283  
 Db 1107 TTTGGTATGCAAGATTTCGAGATCTGCTGCAAAATGCTGGCTCTGCGGCTGTTC 1166  
 QY 284 SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyr 303  
 Db 1167 AGCGACGACTTAATCTGGCTACCTCAATAGCTGCTTCTTAAGGAAGGCCATGTAC 1226  
 QY 304 HisAlaLeuGlnLysAlaArgValArgAlaGlyLysThrPheProSerSerProGlyLys 323  
 Db 1227 CACACTCGAGAAAGCCAGGGTTGAGCTGCAAGACCTTCTCCAGAGACTCTGAGAG 1286  
 QY 324 SerLeuGluAspGlnLeuLysPheMetLeuGlnTyrAlaHisGlyLysPheLysProThr 343  
 Db 1287 TCACCTGGAGGACCAAGCTGAACCCCTGCTGAGTGGGCCCAAGGCTTCAAGCCTACT 1346  
 QY 344 GlyIleGlyLysLeuLysProAsnAsnThrGlnProValAlaAsnLysSerLysValArg 363  
 Db 1347 GGGATCGAGGGCTCTAAACCCAAACAAAGCAACAGTGTAAATCAAGTCAAGGTGGT 1406  
 QY 364 ArgAlaGlySerArgLysLeuGlnSerArgLysTyrGlnAsnLysThrArgArgThr 383  
 Db 407 CGTTCAGACATGAGAACTTGAACCCAGAGACCCGAGAACAAAGTCGAAGACGCCA 1466  
 QY 384 AlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLeuLysThrAsnCys 403  
 Db 1467 ACCAATGACTCTGCTGCTTGAG--TCCCCCCCAACCCGCTTCAAGCAAAATAGC 1523  
 QY 404 TyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGlnMetAlaSerAsp 423  
 Db 1524 TAT--GGCGGGAAGGACCGAGGGAGATGAGAGACCCGAGAACGATGCTTCTAA 1580  
 QY 424 ValAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerCysGlyArgLysAsnPro 443  
 Db 1581 GTCACAAACAAACAGGCAATCTGAGAACCCGCTTTGCTGTGGAAGAGAAACCT 1640  
 QY 444 ValSerPheHisProLeuPheGlnGlyLysCysGlnThrCysArgAspArgPheLeu 463  
 Db 1641 GTGCTCTTCCACCCCTCTTTGAGGGTGGGCTGTCTCAGAGTGCAGGAGTGCCTCTA 1700  
 QY 464 GlnLeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThrValCysGln 483  
 Db 1701 GAGCTCTTCAATGTATGATGAGACCGCTATCAGTCTCTGACCGTGTGCTGTAG 1760  
 QY 484 GlyArgGlnLeuLeuLeuCysSerAsnThrSerCysArgCysPheCysValGlnCys 503  
 Db 1761 GGCGCTGAACCTGCTGTGCAAGTAAACAAAGCTCTGCAAGTGTCTTGTGTGAGGT 1820  
 QY 504 LeuGluValLeuValGlyThrGlyThrAlaAlaGlnAlaLysLeuGlnLupProTyrSer 523

Db 1821 CTGAGAGTGTGCTGGGGGCGACAGCTGAGATGCAAGCTGCAGAAACCTCGAGC 1880  
 QY 524 CysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgLysAspTyrAsn 543  
 Db 1881 TGCATATGTGCTCTCCCTCAGCGCTGCATGGGTCTCTCCGACCCAGAAAGATTGGAC 1940  
 QY 544 ValArgLeuGlnAlaPhePheThrSerAspThrGlyLeu---GlyTyrGlnAlaProLys 562  
 Db 1941 ATGCGCTCGCAAGACTTCTTCACTACTGATCTGACCTCGAAGAAATTTGACCCCAAG 2000  
 QY 563 LeuTyrProAlaIleProAlaAlaArgArgProIleArgValIleuSerLeuPheAsp 582  
 Db 2001 TTGTACCCAGCAATTCCTGCAAGCCAAAGAGGCCATTAAAGTCTGTCTCTTTGAT 2060  
 QY 583 GlyIleAlaThrGlyTyrLeuValLeuLysGlnLeuGlyLysValGlyLysTyrVal 602  
 Db 2061 GGAATTGCAACGGGCTACTGTGCTCTCAGAGATGGGTATTAAGTGAAGAAATCAATT 2120  
 QY 603 AlaSerGluValCysGlnGluSerIleAlaValGlyThrValLysHisGlyLysAsnIle 622  
 Db 2121 GCCTCCGAAGCTGTGAGAGACTCATCGCTGTGGAACTGTAAAGCATGAAGCCAGATC 2180  
 QY 623 LysTyrValAsnAspValArgAsnIleThrLysLysAsnIleGlnGluTyrGlyProPhe 642  
 Db 2181 AAATATGTCATGACGTCCGGAATAATCACCAGAAATAATGGAAGATGGGGCCGTTT 2240  
 QY 643 AspLeuValIleGlyLysSerProCysAsnAspLeuSerAsnValAsnProAlaArgLys 662  
 Db 2241 GACTTGTGATTTGGTGAAGCCATGCAATGATCTCTTAACGTCAATCTGCCCCGAAA 2300  
 QY 663 GlyLeuTyrGlnGlyThrGlyArgLeuPheGlnPheTyrHisLeuLeuAsnTyrSer 682  
 Db 2301 GGTATATATAGGGGCAACAGAAAGCTCTTCTTCAGATTTTAAACCTTGCTGAATTAACC 2360  
 QY 683 ArgProLysGlnGlyAspAspArgProPhePheTyrPheGlnAsnValAlaIleMet 702  
 Db 2361 CGCCCAAGAGGGCGCAACCCGTCAATCTTCTGATGTTGCAAAATGTTGTGGCCATG 2420  
 QY 703 LysValGlyAspLysArgAspIleSerArgPheLeuGlnCysAsnProValIleAsp 722  
 Db 2421 AAAGTGAATGCAAGAAAGATCTCAAGATCTCTGGCATGTAAACCAAGTATGATTCAT 2480  
 QY 723 AlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMet 742  
 Db 2481 GCATCAAGTGTCTGCTGCTCAAGGGCCCGGTACTTCTGGGTAACTTACCCGGAATG 2540  
 QY 743 AsnArgProValIleAlaSerLysAsnAspLysLeuGlnLeuGlnAspCysLeuGlnTyr 762  
 Db 2541 AACAGGCCCGTATGCTTCAAAAGATGATTAAGCTCGAGCTGCAGACTGCTGGAAGTTC 2600  
 QY 763 AsnArgIleAlaLysLeuLysValGlnThrIleThrLysSerAsnSerIleLys 782  
 Db 2601 AGTGGACAGCAAAAGTTAAAGAAAGTGCAGCAATTAACCAACCAAGTCAATCATCA 2660  
 QY 783 GlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysGluAspValLeuTyrCys 802  
 Db 2661 CAGGCGAAACCAAGCTTTTCTCTGTATGATCATGAATGCAAGAGCAGACTTTTGTGGTC 2720  
 QY 803 ThrGlnLeuGlnArgIlePheGlyPheProValHisTyrThrAspValSerAsnMetGly 822  
 Db 2721 ACTGAGCTCGAAAGATCTTGGGCTTCCCTGCTCAGTACAGAGCGTGTCCAAATGGGC 2780  
 QY 823 ArgGlyAlaArgGlnLysLeuGlnLysArgSerTyrProValIleArgHisLeu 842  
 Db 2781 CGCGCGCGCCCTCAAGAGCTGCTGGCAGGTCCTGAGGTATACCGGTATCAGACACCTG 2840  
 QY 843 PheAlaProLeuLysAspTyrPheAlaCysGln 853  
 Db 2841 TTTGCCCTTGAAGGACTTACTTGTCTGTCAA 2873  
 RESULT 7  
 AF151969

LOCUS AF151969 4338 bp mRNA linear ROD 08-JUN-2000  
DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 1  
(Dnmt3b) mRNA, complete cds.  
ACCESSION AF151969  
VERSION AF151969.1 GI:8347117  
KEYWORDS Mus musculus.  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 4338)  
2 (bases 1 to 4338)  
Yin.B., Chen.Y.T., Zhu.M., Luo.Y.J., Zhu.N., Xu.S.C., Wu.G.Y. and  
Shen.Y.  
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing  
isoforms in mouse embryonic tissue  
JOURNAL Unpublished  
REFERENCE 1 (bases 1 to 4338)  
AUTHORS Yin.B., Chen.Y.T., Zhu.M., Luo.Y.J., Zhu.N., Xu.S.C., Wu.G.Y. and  
Shen.Y.  
TITLE Direct Submission  
SUBMITTED (17-MAY-1999) Department of Biochemistry and Molecular  
Biology, Institution of Basic Medical Sciences, Chinese Academy of  
Medical Sciences, Dong San Tiao 5, Beijing City 100005,  
P.R.China  
LOCATION/Qualifiers  
1. 4338  
/organism="Mus musculus"  
/strain="KM"  
/db\_xref="taxon:10090"  
/dev\_stages="8-9 day old embryo"  
1..4338  
/gene="Dnmt3b"  
412..2991  
/gene="Dnmt3b"  
/functions="methylates cytosine in DNA"  
/note="alternatively spliced"  
/codon\_start=1  
/product="DNA cytosine-specific methyltransferase isoform  
1"

## FEATURES

source  
1. 4338  
/organism="Mus musculus"  
/strain="KM"  
/db\_xref="taxon:10090"  
/dev\_stages="8-9 day old embryo"  
1..4338  
/gene="Dnmt3b"  
412..2991  
/gene="Dnmt3b"  
/functions="methylates cytosine in DNA"  
/note="alternatively spliced"  
/codon\_start=1  
/product="DNA cytosine-specific methyltransferase isoform  
1"

## gene

## CDS

1  
/protein\_id="AA074515.1"  
/db\_xref="GI:8347118"  
/translation="MKGDSRLHNEEGASVEECIIIVNGNFSQSSDTPKADSPVPLE  
AICTEPTCTRTGRSSRLSKREVSLLNYTODMTDGDGDDDEVDGNGSDILMPK  
LRTETKTRTRSESAPVTRHNSGTSSLERORASPRITRGORHVVQYVPEVPATR  
SRRRASSASTPWSPASVDFMEBVTPKSVTPSDLSODQEGMDTQVDFGSDIF  
GDSTEYDDKFEFGIDLVNKGKIGFVSWPAMVSVKATSKRQAMPKRWVDFGSDIF  
SEIADKLVALGLFSDHFNATFNKLVSRKMYHTLEKARVAGTFTSSPESLED  
DSASSEPPPKLKTNSYGKDGDESESRMASSEVTNNKLEDRCLSCGKNPV  
PHPLFEGGLQSCSDRLLELYWYDEDEYQYCTVCEGRELILCSNTSCRCRPEC  
LEVYAGTADAKLEPWSCHLFCRGVLRARRKWNRLQDFTTDPPLESFEP  
PKLPAPAKRRPRLVSLDFGATIGLVLELGIKVEKIIASEVCAEISAVGVNKH  
EQIKYVNDVRKTKNIEWGPFDLVIGGSPCNDLSNVNPKRGLEYEGTGRFFEFY  
HLLNTPKSDNRPFPMFNVAMKNDKDIISRLACNPPMIDAIKVSAAHARY  
FWGLPQNNRPVWASKNKLEQDCLFESRTAKLKKVOTITTKNSIRIGKQKLPVY  
MNGKDDVLWCTELERIFGPAHYTIVSNNGRQKLLGRSWSVPVIRHLFAPLKDVF  
ACE"

BASE COUNT 1073 a 1124 c 1173 g 968 t

## ORIGIN

## Alignment Scores:

Pred. No.: 1.73e-190 Length: 4338  
Score: 3748.00 Matches: 716  
Percent Similarity: 88.52% Conservative: 55  
Best Local Similarity: 82.20% Mismatches: 70  
Query Match: 82.08% Indels: 30  
DB: 10 Gaps: 10

US-09-720-086-8 (1-853) x AF151969 (1-4338)

Qy 1 MetLysGlyAspThrArgHisLeuAsnGlyGluGluAspAlaGlyArgGluAspSer 20

Db 412 ATGAAGGAGACAGCAGACACTCTGAATGAAGAGAGGGTGCCAGCGGGTATGAGGAGTGC 471

Qy 21 IleLeuValAsnGlyAlaCysSerAspGlnSerSerAsp-----SerPro 35  
Db 472 ATTATCGTTAATGGAACTTTCAGTGACCACTCTCAGACAGGAAGGATGCTCCCTCACCC 531

Qy 36 ProIleLeuGluAlaIle-----ArgThrProGluIleArgGlyValGAG 50  
Db 532 CCAGTCTTTGGAGGCAATCTCCACAGAGCCAGTCTGCACACAGAGACAGAGCCGAGG 591

Qy 51 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuSerTyThrGlnAspLeu 70  
Db 592 TCACGCTCCCGCTGTCTAAGAGGAGGTCTCCAGCTTCTGAATATACAGCAGGACATG 651

Qy 71 ThrGlyAspGlyAspGly-----GluAspGlyAspGlySerAspThrProVal 86  
Db 652 ACAGGAGATGGAGACAGAGATGATGAATGATGGATGCTCTGATATT---CTA 708

Qy 87 MetProLysLeuPheArgGlu-----ThrArgThrArgSerGluSerProAlaVal 103  
Db 709 ATGCCAAAGCTCACCCCTGAGACCAAGGACACAGGACCGCTCTGAAGCCCGCTGC 768

Qy 104 ArgThrArgAsnAsnSerValSerSerArgGluArgHisArgProSerProArgSer 123  
Db 769 CGAACCCGACATAGCAATGGGACCTCCAGCTTGGAGAGCAAGAGCCTCCCCCAATC 828

Qy 124 ThrArgGlyArgGlnGlyArgAsnHisValAspGluSerProValGluPheProAlaThr 143  
Db 829 ACCCGAGGTGGAGAGGCGCCACCATGTGCAGAGGTACCTCTGGAGTTTCCGGGTACC 888

Qy 144 ArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProThrProSerProProSer 163  
Db 889 AGGTCTCGGAGACGTGAGCATCGTCTTCAGCAAGCAGCGCATGTCATCCCTGCCAGC 948

Qy 164 SerTyrLeuThrIleAspLeuThrAspThrGluAspThrHisGlyThrProGlnSer 183  
Db 949 -----GTCCGACTTCATGGAAGAAGTG-----ACACCTAAGAGC 981

Qy 184 SerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMetGluSerPro 203  
Db 982 GTCAGTACCCCATCAGTTGACTTGAGCCAGGATGGAGATCAGGAGGTATGATACCACA 1041

Qy 204 GlnValGluAlaAspSerGlyAspGlyAspSerGluTyrGlnAspGlyLysGluPhe 223  
Db 1042 CAGGTGGATGACAGAGACATATATGAGACAGACAGCATATCAGGATGATAAAGAGTTT 1101

Qy 224 GlyIleGlyAspLeuValTppGlyLysIleLysGlyPheSerTppTppProAlaMetVal 243  
Db 1102 GGAATAGGTGACCTCGTGGGGAAGATCAAGGGCTTCTCTGGTGGCTCGCATGGTG 1161

Qy 244 ValSerTppLysAlaThrSerLysArgGlnAlaMetSerGlyMetArgTppValGlnTpp 263  
Db 1162 GTGCTCTGGAAGCCACTCAAGCCAGGCGCATGCCGGAATGCCGTGGGTACAGTGG 1221

Qy 264 PheGlyAspGlyLysPheSerGluValSerAlaAspLysLeuValAlaLeuGlyLeuPhe 283  
Db 1222 TTTGGTATGGCAAGTTTCTGAGATCTCTGCTGACAACTGGTGGCTCTGGGGCTGTTC 1281

Qy 284 SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyr 303  
Db 1282 AGCCAGCACTTTAATCTGGCTACCTTCAATAAGCTGGTTTCTTATAGGAAGGCATGTAC 1341

Qy 304 HisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSerProGlyAsp 323  
Db 1342 CACACTCTGGAGAAAGCCAGGGTTGAGCTGCAAGACCTTCTCCAGCAGTCTCGGAGAG 1401

Qy 324 SerLeuGluAspGlnLeuLysProMetLeuGluTppAlaHisGlyLysPheLysProThr 343  
Db 1402 TCACCTGGAGGACAGCTGAAGCCCATGCTGGAGTGGGCCCAAGGGTCTCAAGCCTACT 1461

Qy 344 GlyIleGluGlyLeuLysProAsnAsnThrGlnProValValAsnLysSerLysValArg 363  
Db 1462 GGGATCGAGGGCCCTCAAAACCCAAAGAACCAAGCAGGTGGTTAATAAGTCGAAGGTGCGT 1521

QY 364 ArgAlaGlySerArgGlyLeuGluSerArgGlySerGlyLeuValSerThrArgArgThr 383  
 Db 1522 CGTTCAGACAGTAGGAACTTGGAAACCCAGAGACCGCAAAAGTTCGAAAGCACCA 1581  
 QY 384 AlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLeuLysThrAsnGly 403  
 Db 1582 ACCAATGACTGCTGCTGCTTCTGAG--TCCCCCAACCCAGCCGCTCAAGCAAAATAGC 1638  
 QY 404 TyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluGlnMetAlaSerAsp 423  
 Db 1639 TAT--GGCGGGAAGGACCGAGGGAGATGAGAGACCCGAGAACGATGGCTTCTGA 1695  
 QY 424 ValAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerCysGlyArgLysAsnPro 443  
 Db 1696 GTCACCAACCAAGAGGCAATCTGGAAACCGCTGTTGCTCTGGAAGAAAGAACCTT 1755  
 QY 444 ValSerPheHisProLeuPheGluGlyLysLeuGlySerGlyLeuValSerArgAspArgPheLeu 463  
 Db 1756 GTGTCCTTCCACCCCTTTGAGGGTGGGCTGTGTCAAGATGCGGGATGCTTCTTA 1815  
 QY 464 GluLeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThrValCysCysGlu 483  
 Db 1816 GAGCTCTTCAATGATGATGAGAGACGGCTATCACTCACTGACACCGTGTGCTGTAG 1875  
 QY 484 GlyArgGluLeuLeuLeuCysSerAsnThrSerCysArgCysPheCysValGluCys 503  
 Db 1876 GCGCGTGAACCTGCTGTGTGAGTAAACAAAGCTGCTGACAGTCTTCTGTGTGAGAGT 1935  
 QY 504 LeuGluValLeuValGlyThrGlyThrAlaAlaGluValLeuGluGlnProThrSer 523  
 Db 1936 CTGGAGGTGCTGGTGGGGCGACAGCTGAGATGCGAAGCTGCAAGAACCCGAGAC 1995  
 QY 524 CysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgArgLysAspTyrAsn 543  
 Db 1996 TGCTATATGCTCCTCCCTCAGGGCTGATGGGCTCCCGACGAGGAAGATTGGAAAC 2055  
 QY 544 ValArgLeuGlnAlaPhePheThrSerAspThrGlyLeu--GluTyrGluAlaProLys 562  
 Db 2056 ATGGCGCTCGAAGACTTCTTCACTACTGACTGCTGCAAGAAATATTGAGCACCCCAAG 2115  
 QY 563 LeuTyrProAlaIleProAlaAlaArgArgArgProIleArgValLeuSerLeuPheAsp 582  
 Db 2116 TTGATCCAGAAATTCCTCGACGCCAAAGAGGCCCATTAAGTCTCTCTGTTTAT 2175  
 QY 583 GlyIleAlaThrGlyTyrLeuValLeuLysGluLeuGlyIleLysValGlyLysTyrVal 602  
 Db 1176 CGAATGCAACGGGTAATTGGTGTCTCAAGAGTTGGGATTAAGTGAAGAAAGTACTT 2235  
 QY 603 AlaSerGluValCysGluGluSerIleAlaValGlyThrValLysHisGluGlyAsnIle 622  
 Db 2236 GCGTCCGAAGCTGTGCGAGAGTCCATCGCTGTGGAACTGTTAAGCATGAAGGCCAGATC 2295  
 QY 623 LysTyrValAsnAspValArgAsnIleThrLysLysAsnIleGluGluTTPGlyProPhe 642  
 Db 2296 AAAATATGCAATGACGTCGCCGAAATATCCAAAGAAATATTAAGATGGGGCCGCTTC 2355  
 QY 643 AspLeuValIleGlyLysSerProCysAsnAspLeuSerAsnValAsnProAlaArgLys 662  
 Db 2356 GACTTGGGATTTGGTGGAAACCCATGCAATGATCTCTTAACGTCATACCTGCCGCAAA 2415  
 QY 663 GlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPheTyrHisLeuLeuAsnTyrSer 682  
 Db 2416 GGTATATATGAGGACACGAAAGGCTTCTTCCGATTTTACCACTTGTCTGATTAATAC 2475  
 QY 683 ArgProLysGluGlyAspAspArgProPhePheTyrPheTyrPheGluAsnValAlaIleMet 702  
 Db 2476 CGCCCAAGAGGGGCGAACAACGTCATCTTCTGAGTTCGAAAGTTCGAGAAATGTGTGCCATG 2535  
 QY 703 LysValGlyAspLysArgAspIleSerArgPheLeuGluCysAsnProValMetIleAsp 722  
 Db 2536 AAAGTGAATGACAAAGAAACATCTCAAGATTCCTGGCATGTAAACCCAGTATGATCAT 2595  
 QY 723 AlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMet 742

Db 2596 GCCATCAAGGAGTGTCTGCTGCTCACAGGCCCGGATCTTCTGGGTAACTCCCGGAATG 2655  
 QY 743 AsnArgProValIleIleAlaSerLysAsnAspLysLeuGluGlnAspCysLeuGluTyr 762  
 Db 2656 AACAGCCCGGTGATGGCTTCAAGAAATGATAGCTCGAGCTGCAAGATGCTGGAGTTC 2715  
 QY 763 AsnArgIleAlaLysLeuLysValGlnThrIleThrThrLysSerAsnSerIleLys 782  
 Db 2716 AGTAGACAGCAAAAGTTAAAGAAATGACAGCAATTAACCAAGTCAATCTCCATCGA 2775  
 QY 783 GlnGlyLysAsnGlnLeuPheProValAlaMetAsnGlyLysGluAspValLeuTyrCys 802  
 Db 2776 CAGGCAAAAACCAAGCTTTTCTCTGTAGTCAATGAGCAAGACGACGATTTGTGTGC 2835  
 QY 803 ThrGluLeuGluArgIlePheGlyPheProValHisTyrThrAspValSerAsnMetGly 822  
 Db 2836 ACTAGCTCGAAAGATCTTCCGCTTCCCTGCTCACTACAGAGAGTGTCCAAACATGGGC 2895  
 QY 823 ArgGlyAlaArgGlnLysLeuLeuGlyArgSerTyrSerValProValIleArgHisLeu 842  
 Db 2896 CGCGGCGCCCTCAGAACTGCTGGCAGGTCCTGAGTGTACCGGTATACAGACACTG 2955  
 QY 843 PheAlaProLeuLysAspTyrPheAlaCysGlu 853  
 Db 2956 TTGCCCCCTTGAAGAGACTACTTGGCTGTGA 2988  
 RESULT 8  
 AF068627 4135 bp mRNA linear ROD 06-DEC-1999  
 LOCUS Mus musculus DNA cytosine-5 methyltransferase 3B2 (Dnmt3b) mRNA, 2988 bp.  
 DEFINITION Mus musculus DNA cytosine-5 methyltransferase 3B2 (Dnmt3b) mRNA, alternatively spliced, complete cds.  
 ACCESSION AF068627  
 VERSION AF068627.2 GI:6449471  
 KEYWORDS  
 SOURCE  
 ORGANISM Mus musculus.  
 Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 4135)  
 Okano, M., Xie, S., and Li, E.  
 Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases  
 Nat. Genet. 19 (3), 219-220 (1998)  
 JOURNAL MEDLINE 98324766  
 PUBMED 9662389  
 REFERENCE 2 (bases 1 to 4135)  
 Xie, S., Okano, M., and Li, E.  
 Direct Submission  
 Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street, Charlestown, MA 02129, USA  
 JOURNAL 3 (bases 1 to 4135)  
 Okano, M., Chijiwa, T., Sasaki, H., and Li, E.  
 Direct Submission  
 Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street, Charlestown, MA 02129, USA  
 REMARK Sequence update by submitter  
 On Nov 18, 1999 this sequence version replaced gi:3527981.  
 FEATURES  
 source  
 1. 4135  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="2"  
 /map="84.0 cM"  
 gene  
 1. 4135  
 /gene="Dnmt3b"  
 misc\_feature  
 1. 4135  
 /gene="Dnmt3b"  
 /note="similar to EST sequences deposited in GenBank Accession Numbers AA116694, AA119979, AA177277, AA210568, AA407106, and AA575617"  
 CDS  
 269..2788  
 /gene="Dnmt3b"



```

/functions="de novo DNA methylation"
/notes="alternatively spliced product; contains Cys-rich
region; C-terminal region is similar to corresponding
region of DNA cytosine-5 methyltransferase 3A"
/codon_start=1
/product="DNA cytosine-5 methyltransferase 3B2"
/db_xref="GI.6449472"
/translation="MKGDSHLENEEGAGYVEGCIIVNGNFSDOSSPTKDAAPVLE
LITREPTVCTETRGSRSSRLSKREVSSLLAYQMDTGDDRDDEVDGNGSDILMPK
LITREKTRIRSESPAVRTHNSGTSISLERORASFRIRGRQGRHVRQIFVEFPATR
SKRRSSASSTPWSPASVDEMEVTPKSVTFSDLSQDQGGMDITQVDAESRD
SETEYQDDKEFGIGLVKMGKIFGFWMPVAVSWKATSKRQAMPGRWQFGDGKF
SIEADKLVALGFLGSQHFNLATFNKLVSRKAMVHTLEKARVRAGKTFFSSPGESELD
QKPMLEWAGHGKPTGIEGLPKKKPENKRRRTTNDASAESPPLKRLKTNISYGG
KRDGEDSESRMASVETNNKGNLDRCLSCGKNPVSPHPLFEGGLCQSCDRLEL
FYMDPEQSYCTVCCGERELKLSNTSCRCFCVECLVVGAGTAAEDAKLQEPWS
CYMLQPCORGLVRRKRDNMRLQDFTTDPDLEFPFKLPAIIPAAKRRIRVLISL
FDGATLVGLVKBELGKVEKYIASEVCAESIAGVTVKHEGQIKYVNDVRKTIYKNIEE
WGPFDLVIGSPCNDLSNVNPAKGLYEGTGLRFFBYHLLNLTTPKRGEDNRPFFWMPF
ENVVAMKVNKKDQISRLACNPNMIDAIKVSAAHARYFWGNLPNGNRPVMAKNDKL
ELQDLEFSRTAKLKKVQITTTKSNISIRQGNQLFPVVMNGKDDVLWCTELERIFGFP
AHYTDVSNMGRGAROKLLGRSWSVPVIRHLEAPLKYFACE"
BASE COUNT 1043 a 1048 c 1106 g 938 t
ORIGIN

```

## Alignment Scores:

```

Pred. No.: 1,25e-186 Length: 4135
Score: 1675.00 Matches: 703
Percent Similarity: 86.80% Conservative: 53
Best Local Similarity: 80.71% Mismatches: 65
Query Match: 80.49% Indels: 50
DB: 10 Gaps: 11

```

US-09-720-086-8 (1-853) x AF068627 (1-4135)

Qy	1	MetLeuGlyAspThrArgHisLeuAsnGlyGluGluAspAlaGlyGlyArgGluAspSer	20	Db	806	-----GTCGACTTCATGGAAGAAGTG-----     :	838
Db	269	ATGAAGGGAGACAGACATCTGAATGAAGAAGAGGGTGCAGCGGTATGATGAGATGC	328	Qy	184	SerSerThrProTyrAlaArgLeuAlaGlnAapSerGlnGlnGlyGlyMetGluSerPro	203
Qy	21	IleLeuValAsnGlyAlaCysSerAspGlnSerSerAsp-----SerPro	35	Db	839	GTCACTACCCCATCAGTTGACCTTGAGCCAGGATGGAGATCAGAGGGTATGATACCACA	898
Db	329	ATTATCGTTAAATGGGAACCTTCAGTGACCACTCTCGACACACAGGAGTGTCCCTCACCC	388	Qy	204	GlnValGluAlaAapSerGlyAapGlyAapSerSerGluTyrGlnAapGlyGlyGluPhe	223
Qy	36	ProIleLeuGluAlaIle-----ArgThrProGluIleArgGlyArgArg	50	Db	899	CAGGTGGATGAGACAGACAGATGGACACACAGATATCAGGATGATGAAGAGTTT	958
Db	389	CCAGTCTTGGAGGCAATCTGCACAGAGCCAGTCTGCACACACAGAGACAGAGGCCGAGG	448	Qy	224	GlyIleGlyAapLeuValTrpGlyLysIleLysGlyPheSerTrpTrpProAlaMetVal	243
Qy	51	SerSerSerArgLeuSerLeuArgGluValSerSerLeuLeuSerTyrThrGlnAapLeu	70	Db	959	GAAATAGGTGACCTCGTGTGGGAAAGATCAAGGGCTTCTCCTGGTGGCTGCGCATGGTG	1018
Db	449	TCAAGTCCCGAGTCTTAAGAGGGAGGTCTCCAGCCTTCTCGAATTACACGACGACATG	508	Qy	244	ValSerTrpLysAlaThrSerLysArgGlnAlaMetSerGlyMetArgTrpValGlnTrp	263
Qy	71	ThrGlyAspGlyAspGly-----GluAspGlyAspGlySerAspThrProVal	86	Db	1019	GTGTCTCTGGAAAGCACCTCCAAGCGACAGGCGCATGCCCGGAATGCCGTGGTACAGTGG	1078
Db	509	ACAGGAGATGGAGACAGAGATGATGAAGTAGATGGGATGGCTCTGATATT-----CTA	565	Qy	264	PheGlyAapGlyLysPheSerGluValSerAlaAapLysLeuValAlaLeuGlyLeuPhe	283
Qy	87	MetProLysLeuPheArgGlu-----ThrArgThrArgSerGluSerProAlaVal	103	Db	1079	TTTGGTGATGGCAAGTTTCTGAGATCTCTGCTGACAACTGGTGGCTCTGGGGCTGTTC	1138
Db	566	ATGCCAAAGCTCACCGGTGAGACCAAGGACACGAGTCTCTGAAAGCCCGGTGTC	625	Qy	284	SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyr	303
Qy	104	ArgThrArgAsnAsnSerValSerSerArgGluAlaArgHisArgProSerProArgSer	123	Db	1139	AGCCAGCAGCACTTAACTTGGCTACCTTCAATAAGCTGTTTCTTATAGGAGGCCATGTAC	1198
Db	626	CGAACCCGACATAGCAATGGGACCTCCAGCTTGGAGAGCAAGAGCCCTCCCGCAGATC	685	Qy	304	HisAlaLeuGluLysAlaArgValAlaArgAlaGlyLysThrPheProSerSerProGlyAsp	323
Qy	124	ThrArgGlyArgGlnGlyArgAsnHisValAspGluSerProValGluPheProAlaThr	143	Db	1199	CACACTCTGGAGAAGCCAGGGTTTCAGCTGGCAAGACCTTCTCCAGCAGTCTCTGGAGAG	1258
Db	686	ACCCGAGGTTCGGCGGGCCGACCATGTGTCAGAGAGTACCTCTGTGGAGTTTTCGGGTACC	745	Qy	324	SerLeuGluAapGlnLeuLysProMetLeuGluTrpAlaHisGlyLysPheLysProThr	343
Qy	144	ArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProTrpProSerProProSer	163	Db	1259	TCACTGGAGGACCACTGGAAGCCCATGCTGGAGTGGGGCCACGGTGGCTCTCAAGCCTACT	1318
Db	746	AGGTCTCGGAGACGTCGAGCATCTGCTTCAGCAAGCACGCCCATGTCATCCCTGCCAGC	805	Qy	344	GlyIleGluGlyLeuLysProAsnAsnThrGlnProValValAsnLysSerLysValArg	363
Qy	164	SerTyrLeuThrIleAspLeuThrAspThrGluAspThrHisGlyThrProGlnSer	183	Db	1319	GGGATCGAGGGCCCTCAACCCAAAGAGCAACCA-----GAGAACAAAGTCAAGACGCACA	1378
				Qy	364	ArgAlaGlySerArgLysLeuGluSerArgLysTyrGluAsnLysThrArgArgArgThr	383
				Db	1355	-----GAGAACAAAGTCAAGACGCACA	1378
				Qy	384	AlaAapAspSerAlaThrSerAspTyrCysProAlaProLysArgLysLeuLysThrAsnCys	403
				Db	1379	ACCAATGACTCTGCTGCTTCTGAG---TCCCCCCCCCAAGCGCTTCAAGACAAATAGC	1435
				Qy	404	TyrAsnAsnGlyLysAspArgGlyAapGluAapGlnSerArgGluGlnMetAlaSerAsp	423
				Db	1436	TAT---GGCGGGAAGGACCGAGGGGAGGATGAGGAGAGCGAGAACGGATGGCTTCTGAA	1492
				Qy	424	ValAlaAsnAsnLysSerSerLeuGluAapGlyCysLeuSerCysGlyArgLysAsnPro	443
				Db	1493	GTCCACCAACACAAAGGCAATCTGGAAAGCCGCTGTTTCTCTGTGGAAAGAAAGACCT	1552
				Qy	444	ValSerPheHisProLeuPheGluGlyLeuCysGlnThrCysArgAspArgPheLeu	463
				Db	1553	GTGTCCTTTCACCCCTCTTTGAGGTGGGCTGTGTAGAGTTGCCGGGATCGCTTCTCTA	1612
				Qy	464	GluLeuPheTyrMetTyrAspAapGlyTyrGlnSerTyrCysThrValCysCysGlu	483
				Db	1613	GAGCTCTTCTACATGATGATGAGGACGGCTATCAGTCTCTACTGACCGGTGTGTGTGAG	1672
				Qy	484	GlyArgGluLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCysValGluCys	503
				Db	1673	GGCCGTGAACCTGCTGTGTGTCAGTAACAAGCTGCTGCAGATGCTTCTGTGTGGAGTGT	1732
				Qy	504	LeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnGluProTrpSer	523
				Db	1733	CTGGAGGTGCTGTGTGGCGCAGGCACAGCTGAGGATGCCAAGCTGCAAGAACCTTGGAGC	1792
				Qy	524	CysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgArgLysAspTrpAsn	543



Db 1793 TGCTATATGTCCTCCCTCAGCGCTGCATGGGCTCTCCGACGAGAAAGATTGGAAC 1852  
Qy 544 ValArgLeuGlnAlaPhePheThrSerAspThrGlyLeu---GlyTyrGlnAlaProlys 562  
Db 1853 ATGGCGCTCGAAGACTTCTTCACTGATCTGTCGAGAAAGATTGAGCCACCCAG 1912  
Qy 563 LeuTyrProAlaIleProAlaAlaArgArgProIleArgValLeuSerLeuPheAsp 582  
Db 1913 TTGTACCCAGAAATTCCTCGCAGCAAAAGAGGCCATTAAAGTCTGCTCTGTTTAT 1972  
Qy 583 GilylIleAlaThrGlyTyrLeuValLeuGlyGlyIleLysValGlyLysTyrVal 602  
Db 1973 GGAATTGCAAGCGGCTACTTGTGCTCAGAGAGTTGGGATTTAAAGTGGAAAGTACTT 2032  
Qy 603 AlaSerGlnValCysGlnGluSerIleAlaValGlyThrValLysHisGluGlyAsnIle 622  
Db 2033 GCCTCCGAGCTGTGCGAGACTGCTGTGGGAACTGTTAAGCATGAAGGCGAGATC 2092  
Qy 623 LysTyrValAsnAspValArgAsnIleThrLysLysAsnIleGluGluTyrGlyProPhe 642  
Db 2093 AAATATGTCATGACGTCCGGAATAATCACAAGAAATATGAAAGTGGGCGCCCTTC 2152  
Qy 643 AspleuValIleGlyLysSerProCysAsnAspLeuSerAsnValAsnProAlaArgLys 662  
Db 2153 GACTTGGATGGTGGAGGCCCATGATATATCTCTTAAGCTCAATCTGCGCCGAA 2212  
Qy 663 GlyLeuTyrGlnGlyTyrGlyArgLeuPheGluPheGlyHisLeuLeuAsnTyrSer 682  
Db 2213 GCTTTATATGAGGCGACAGAAAGGCTCTTCCAGTTTACACTTCTGATTTATAC 2272  
Qy 683 ArgProLysGlnGlyAspAspArgProPhePheThrMetPheGluAsnValValAlaMet 702  
Db 2273 CGCCCAAGGAGGCGAGCAACCGTTCATTCTTCGAGTGTGCAAAATGTTGTGGCCATG 2332  
Qy 703 LysValGlyAspLysArgAspIleSerArgPheLeuGluCysAsnProValMetIleAsp 722  
Db 2333 AAAGTGAATGCAAGAAAGACATCTCAAGATTCTCGGATGTAACCCAGTATGATCGAT 2392  
Qy 723 AlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTyrPGLYAsnLeuProGlyMet 742  
Db 2393 GCCATCAAGGTGTCTGCTGCTCAGAGGCCCGGTACTTCTGGGGTAACCTACCCGGAATG 2452  
Qy 743 AsnArgProValIleAlaSerLysAsnAspLysLeuGlnLeuGlnAspCysLeuGluTyr 762  
Db 2453 AACAGCGCCCGATGAGCTTCAAGAAATGATAGCTCGAGTCGACAGACTGCTGGAGTTTC 2512  
Qy 763 AsnArgIleAlaLysLeuLysValGlnThrIleThrThrLysSerAsnSerIleLys 782  
Db 2513 AGTAGGACAGCAAGTTTAAAGAAAGTGCAGCAATTAACCCAGTCAACTCCATCAGA 2572  
Qy 783 GlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysGluAspValLeuTyrCys 802  
Db 2573 CAGGGCAAAAACCACTTTTCCCTGTACTGATGCAAGAGAGCAAGCTTTTGTGGTGC 2632  
Qy 803 ThrGluLeuGlnArgIlePheGlyPheProValHisTyrThrAspValSerAsnMetGly 822  
Db 2633 ACTGAGCTCGAAGAGATCTTGGGCTTCCCTGCTCACTACAGCAGCTGCCAACAATGGGC 2692  
Qy 823 ArgGlyAlaArgGlnLysLeuLeuGlnLysArgSerTyrSerValProValIleArgHisLeu 842  
Db 2693 CGCGCGCGCGCTCAGAAAGCTGTGGGAGGTCTTGAGTGAACCGGTATCAGACACCTG 2752  
Qy 843 PheAlaProLeuLysAspTyrPheAlaCysGlu 853  
Db 2753 TTTGGCCCTTGAAGGACTACTTGTCTGTGAA 2785  
RESULT 9  
AF151974  
LOCUS AF151974  
DEFINITION Mus musculus DNA cytosine--specific methyltransferase isoform 6  
(Dnmt3b) mRNA, complete cds.  
ACCESSION AF151974  
VERSION AF151974.1 GI:8347130

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS  
FEATURES  
source  
gene  
CDS  
Location/Qualifiers  
1. .4163  
/organism="Mus musculus"  
/strain="KM"  
/db\_xref="taxon:10090"  
/dev\_stage="8-9 day old embryo"  
1. .4163  
/gene="Dnmt3b"  
297. .2816  
/gene="Dnmt3b"  
/functions="methylates cytosine in DNA"  
/note="alternatively spliced"  
/codon\_start=1  
/product="DNA cytosine-specific methyltransferase isoform 6"  
/protein\_id="AA574530.1"  
/db\_xref="GI:8347131"  
/translation="MKGDSRLNEEGAGSYEECIIVNGNFSQSSDTKQAPSPVILE  
ALCTEPCPTETGRSSSRSLSRKREVSILANTYQDMTGDGRDDEVDGSDILMPK  
LRETECDTRPSESAPVATPHSNGTSLEIQRASPRTRRQGRHVEYVPVFPATP  
RRRRASASASPTWSSPASVDPMEEVTPKRSVSPVDSODGOEGMDTTOVDAESIY  
GDSDEVODDPERIGDLYNKIKTGFSPWMPAMVYSWKATSRQAMPGRWOMFGDGKE  
SETISADLVALGLPSGFNLATFNKLVTSTRKMYTHLEKARVAGKTFSSPGESLED  
QLKPMLEMAHGFPTGIEGLKPNKKOPEKRSRRRTNDSASBPPLKLTNYSYG  
KDRGDEBESRERMASEVTNNKGNLEDRCLSCGKKNVSPFLPEGGLCOGRFLEL  
FYVYDEGYSCTVCCGEHELICSNSTSCRCFCEVCELVLAGAGDAEAKLQEPWS  
CVMCLPQRCHGVRRRDKMMRLDPEFTDPDLEEPPEPLLYALIPAARRPTRVLSL  
FDGIATGYLVKELGKIKYKTIASVCABESIANGYKHBOIKYNDVRKITKNIEE  
WGPFVLTGSPENDLSNVAPARKGLYEGTGLFEFYHLNLTREKGNRPFFMF  
ENVVAKVNDKDIISFELACNPVMDAIKYSAHARAYFNGNLGNMRPVMAKNDKL  
ELDDCLEFSRTAKLKVKVQITTKNSNIRQGNOLFVYVNGKXDVLCETELERIFGPP  
AHYTDVSNMGRAROKLGRSWSVPVIRHLFAPLKDYFACE"  
BASE COUNT 1032 a 1074 c 1128 g 929 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 4.85e-186 Length: 4163  
Score: 3664.00 Matches: 702  
Percent Similarity: 86.68% Conservative: 53  
Best Local Similarity: 80.60% Mismatches: 66  
Query Match: 80.25% Indels: 50  
Gaps: 11  
US-09-720-086-8 (1-853) x AF151974 (1-4163)  
Qy 1 MetLysGlyAspThrArgHisLeuAsnGlyGlnGluAspAlaGlyLysArgGluAspSer 20  
Db 297 ATGAAGGAGACACAGACATCTGAATGAAGAAGAGGTGCGACGGGTATAGAGAGCGC 356  
Qy 21 IleLeuValAsnGlyAlaCysSerAspGlnSerSerAsp-----SerPro 35  
Db 357 ATTATCGTTAATGGGAAGTCTCAGTGCACAGTCTCAGACAGAAAGATGCTCCCTCAGCC 416  
Qy 36 ProIleLeuGlnAlaIle-----ArgThrProGluIleArgGlyArgArg 50



QY 763 AsnArg11ealalyLeuLyelyleValGlnThrIleThrThryleSerAsnSerIleIys 782  
:::|||||  
Db 2541 AGTAGGACAGCAAGTAAAGAAAGTGCAGACAAATACCAACCAAGTGCAGATCCATCAGAGA 2600  
QY 783 GlnIlylsAsnGlnleuPheProValValMetAsnGlylySGIuAspValleuTrpCys 802  
|||||CAGGGCAAAAACCAAGCTTTTCCCTGAGTCATGAGTGGCAAGAGCTTTTGTGTGGTC 2660  
Db 2601 CAGGGCAAAAACCAAGCTTTTCCCTGAGTCATGAGTGGCAAGAGCTTTTGTGTGGTC 2660  
QY 803 ThrGluLeuGluAlaGlyIlePheGlyPheProValIleTyrThrAspValSerAsnMetGly 822  
Db 2661 ACTAGAGTCGAAAGATCTTGGCTTCCCTCTCCTACAGAGAGCTGTCCAAATGGGC 2720  
QY 823 ArgGlyAlaArgGlnLyLeuLeuGlyArgSerTrpSerValProValIleArgHisLeu 842  
Db 2721 CGCGCGCCCGCTCAGAAAGCTGCTGGGCGAGTCTGAGTACCGGTCAATCAGACACCTG 2780  
QY 43 PheAlaProLeuLyAspTyrPheAlaCysGlu 853  
|||||TTTGCCCCCTTGAAGACTACTTGTCCGTGAA 2813  
Db 2781 TTTGCCCCCTTGAAGACTACTTGTCCGTGAA 2813  
RESULT 10  
AF151970  
LOCUS AF151970 4278 bp mRNA linear ROD 08-JUN-2000  
DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 2  
ACCESSION AF151970  
VERSION AF151970  
KEYWORDS AF151970.1 GI:8347119  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 4278)  
Yin, B., Chen, Y.T., Zhu, M., Luo, Y.J., Zhu, N., Xu, S.C., Wu, G.Y. and  
Shen, Y.  
Cloning of full-length Dnmt3b cDNA and its alternative splicing  
isoforms in mouse embryonic tissue  
Unpublished  
2 (bases 1 to 4278)  
Yin, B., Chen, Y.T., Zhu, M., Luo, Y.J., Zhu, N., Xu, S.C., Wu, G.Y. and  
Shen, Y.  
Direct Submission  
Submitted (17-MAY-1999) Department of Biochemistry and Molecular  
Biology, Institution of Basic Medical Sciences, Chinese Academy of  
Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,  
P.R.China  
FEATURES  
source  
Location/Qualifiers  
1..4278  
/organism="Mus musculus"  
/strain="KM"  
/db\_xref="taxon:10090"  
/dev\_stages="8-9 day old embryo"  
1..4278  
/gene="Dnmt3b"  
412..2931  
/name="Dnmt3b"  
/function="methylates cytosine in DNA"  
/note="alternatively spliced"  
/codon\_start=1  
/product="DNA cytosine-specific methyltransferase isoform  
2"  
/protein\_id="AAF74516.1"  
/db\_xref="GI:8347120"  
/translation="MKGDSRHUNEEBAGSYECETLYNGNFSDDGSDTAKAPPEVLE  
ALICEPVCTPETRGRSSSRUSKREVSILNYTQDMTGDDRDDEVDDGSGDILMK  
LIREYKDTIRSRSEPAVTRHSNGTSSLEQRPRTIRGOGHHVOEYVEFEPAR  
SRRRASSASATPWSASVADPMEVTPKSTPSVLSODGDEGADTQVDESELY  
GDSREYDODKEFGIDLVMGKIKGFSPMVPVSVKATSRQAMPGKRWOMQEDGKE  
SEISADKLVALGFSQHPNLAPEKIVSYKAWYHTLEKARVAPKGTPTSSRGSLED  
QAKMEWMAHGKFPPTGIBGKPKKQPKNSRRRTTNDASAPPKKLTNYSYG  
KORGEDESESRMASEVTNNKGNLEDCLSGKKNPVSPFLFEGGLCQSCRFLEL  
FYWYDEGYSYCTVCCGEBLLGNSNTSCRCPCVCELEVLVAGTAEDAKLOEPS

Alignment Scores:  
Pred. No.: 5, 01e-186 Length: 4278  
Score: 3664.00 Matches: 702  
Percent Similarity: 86.68% Conservative: 53  
Best Local Similarity: 80.60% Mismatches: 66  
Query Match: 80.25% Indels: 50  
DB: Gaps: 11  
US-09-720-086-8 (1-853) x AF151970 (1-4278)  
QY 1 MetLySGlyAspThrArgHisLeuAsnGlyGluGluAspAlaGlyIleArgGluAspSer 20  
Db 412 ATGAAGGAGACAGCAGACATCTGAATGAAGAAGGGTGCAGCGGTATGAGAGAGTC 471  
QY 21 ILeuValAsnGlyAlaCysSerAspGlnSerSerAsp-----SerPro 35  
Db 472 ATTATCGTTAATGAGAACTTCAGTACAGTCCCAACACGAAAGATGCTCCCTCACCC 531  
QY 36 ProIleLeuGluAlaIle-----ArgThrProGluIleArgGlyIleArg 50  
Db 532 CCACTTGGAGGCAATCTGCACAGACCAGCTGCACACCAAGACCAAGAGCGGAGG 591  
QY 51 SerSerSerArgLeuSerLyArgGluValSerSerLeuLeuSerTyrThrGlnAspLeu 70  
Db 592 TCAAGCTCCCGGCTGTCTTAAGAGGAGGTCTCCAGCTTCTGAATTAACGACGACATG 651  
QY 71 ThrGlyAspGlyAspGly-----GluAspGlyAspGlySerAspThrProVal 86  
Db 652 ACAGAGATGAGACAGACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 708  
QY 87 MetProLyLeuPheArgGlu-----ThrArgThrArgSerGluSerProAlaVal 103  
Db 709 ATGCCAAGCTCACCCCTGAGAACCAAGACACCAAGACACGAGCGCTTGAAAGCGCGGCTGTC 768  
QY 104 ArgThrArgAsnAsnAsnSerValSerSerArgGluArgHisArgProSerProArgSer 123  
Db 769 CGAACCCGACATACCAATGGACCTCCAGCTTGAGAGGCAAAAGAGCTCCCCAGAAATC 828  
QY 124 ThrArgGlyArgGlnGlyArgAsnHisValAspGluSerProValGluPheProAlaThr 143  
Db 829 ACCGAGAGTCCGACAGGCGCGCACCATGTGCAGAGATGACCTGTGGAGATTCCCGGTACC 888  
QY 144 ArgSerLeuArgArgArgAlaThrAlaSerAlaGlyThrProCTProSerProProSer 163  
Db 889 AGGTCTGGAGACCTCAGACATGCTCTTCAGCAAGCAGCGCATGTATCCCTCCAGAC 948  
QY 164 SerTyrLeuThrIleAspLeuThrAspAspThrGluAspThrHisGlyThrProGlnSer 183  
Db 949 -----GTCACTTCATGGAAGAGT-----ACACCTTAAGAGC 981  
QY 184 SerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMetGluSerPro 203  
Db 982 GTCACTACCCCATCAGTTGACTTACAGCAGAGATGAGATGAGAGGATATGATGATACCA 1041  
QY 204 GlnAlaGluAlaAspSerGlyAspGlyAspSerSerGlnTyrGlnAspGlyLyGluPhe 223  
Db 1042 CAGGTGATGAGAGACCATATATATGAGACGACCAAGATATCGAGTAAAGAGTTT 1101  
QY 224 GlyIleGlyAspLeuValITPGLYLSIleGlyIlePheSerTPTPProAlaMetVal 243  
Db 1102 GGAATAGGTAGCTCGTGTGGGAAATCATCAAGGCTTCTCTGTGGCTGCCATGGTG 1161  
QY 244 ValSerTrpLyAlaThrSerLyArgGlnAlaMetSerGlyMetArgTrpValGlnTrp 263

Db	1162	GTGTCCTGGAAAGCCACCTCCAGCGACAGGCCATGCCCGGAATCGCTGGGTACAGTGG	1221
Qy	264	PheGlyAspGlyLysPheSerGluValSerAlaAspLysLeuValAlaLeuGlyLeuPhe	283
Db	1222	TTTGGTGATGGCAAGTTTCTTGAGATCTCTGCTGACAAACTGGTGGCTCTGGGCTGTTT	1281
Qy	284	SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyr	303
Db	1282	AGCCACCACTTTAATCTGGCTACCTTCAATAAGCTGGTTCTTATAGGAAGGCCATGTAC	1341
Qy	304	HisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSerProGlyAsp	323
Db	1342	CACACTCTGGAGAAAGCCAGGGTTCCAGCTGGCGAAGCCTTCTCCAGCAGTCTTGGAGAG	1401
Qy	324	SerLeuGluAspGlnLeuLysProMetLeuGluTyrAlaHisGlyGlyPheLysProThr	343
Db	1402	TCACTGGAGGACCACTGAACCATCTGAGTGGGCCCCAGGGTGGTCTCAAGCCCTACT	1461
Qy	344	GlyIleGluGlyLeuLysProAsnAsnThrGlnProValValAsnLysSerLysValArg	363
Db	1462	GGGATCGAGGGGCTCAAAACCCCAAGAACCAACCA-----1497	
Qy	364	ArgAlaGlySerArgLysLeuGluSerArgLysTyrGluAsnLysThrArgArgThr	383
Db	1498	-----GAGAACAAAGTCGAAGACGACACA1521	
Qy	384	AlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLysThrAsnCys	403
Db	1522	ACCAATGACTCTGCTGCTCTGAG---TCCCCCCCCCAAGCGCTTCAAGACAAATAGC1578	
Qy	404	TyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluMetAlaSerAsp	423
Db	1579	TAT---GGCGGAAGGACCGAGGGGAGGATGGATGGAGAGCGCGAAGACGGATGGCTTCTGAA1635	
Qy	424	ValAlaAsnLysSerSerLeuGluAspGlyCysLeuSerCysGlyArgLysAsnPro	443
Db	1636	GTCAACCAACAAAGGCAATCTGGAAGACCGCTGTTGCTCTCTGGAAAGAACACCT1695	
Qy	444	ValSerPheHisProLeuPheGluGlyGlyLeuCysGlnThrCysArgAspArgPheLeu	463
Db	1696	GTCTCTCTCCACCCCTCTTTGAGGGTGGGCTCTGTCAGAGTTCGGGGATCGGTCCTTA1755	
Qy	464	GluLeuPheTyrMetTyrAspAspGlyTyrGlnSerTyrCysThrValCysGlu	483
Db	1756	GAGCTCTTCTACATGTATGTATGAGGACGGCTATCAGTCTACTGCACCGTGTGTGTAG1815	
Qy	484	GlyArgGluLeuLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnLysProTyrSer	503
Db	1816	GGCGGTGAACCTCTCTCTGTCAGTAACACAAAGCTGCTGCAGATGCTTCTGTGTGAGTGT1875	
Qy	504	LeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnLysProTyrSer	523
Db	1876	CTGGAGTCTCTGTGGCGCCAGCCAGCTGAGATGCCAAGTCCAGTCCAGGAACCTCTGAGC1935	
Qy	524	CysTyrMetCysLeuProGluArgCysHisGlyValLeuArgArgArgLysAspTyrAsn	543
Db	1936	TGCTATATGTGCTCTCCCTCAGCGCTGCGGCTCTCCAGCGAGGAAAGATGGAAC1995	
Qy	544	ValArgLeuGlnAlaPhePheThrSerAspThrGlyLeu---GlyTyrGluAlaProLys	562
Db	1996	ATCGCGCTCCAGAGCTTCTTCACTACTGATCTCTGACCTCGAAGATTTTGAAGCCACCAAG2055	
Qy	563	LeuTyrProAlaIleProAlaAlaArgArgProIleArgValLeuSerLeuPheAsp	582
Db	2056	TTGTACCCAGCAATTCCTGCAGCCAAAGAGGAGGCCCATTAGAGTCTCTCTCTTTGAT2115	
Qy	583	GlyIleAlaThrGlyTyrLeuValLeuLysGluLeuGlyIleLysValGlyLysTyrVal	602
Db	2116	GGAATTCGAACGGGTACTTGGTCTCAAGGAGTTGGGTATTAAAGTGGAAAGTACATT2175	
Qy	603	AlaSerGluValCysGluGluSerIleAlaValGlyThrValLysHisGluGlyAsnIle	622
Db	2176	GCCTCCGAAGTCTGTGCAGAGTCCATCGCTGTGGAACTGTTAAGCATGAAGGCCAGATC2235	
Qy	623	LysTyrValLeuAsnAspValArgAsnIleThrLysLysAsnIleGluGluTyrGlyProPhe	642
Db	2236	AAATATGTGTCATGACGTCCCGAAATTCACCAAGAAAAATATTGAAGAGTGGGCCCGTTTC2295	
Qy	643	AspLeuValIleGlyGlySerProCysAsnAspLeuSerAsnValAsnProAlaArgLys	662
Db	2296	GACTTGGTGATTGGTGGAAAGCCCATGCAATGATCTCTAAACGTCAATCTGCGCCGCAA2355	
Qy	663	GlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPheTyrHisLeuLeuAsnTyrSer	682
Db	2356	GGTTTATATATGAGGCGCACAGGAAGGCTTCTTCGAGTTTTACCACTTGTCTGAATTATACC2415	
Qy	683	ArgProLysGluGlyAspAspArgProPheThrPheMetPheGluAsnValAlaMet	702
Db	2416	CGCCCAAGAGGCGCAACCGCTCCATTCCTTGGATGTTTCGAGATGTTGGCCATG2475	
Qy	703	LysValGlyAspLysArgPheIleSerArgPheLeuGluCysAsnProValMetIleAsp	722
Db	2476	AAAGTGAATGACAAGAAAGACATCTCAAGATTCTTGGCATGTAAACCCAGTGATGATCGAT2535	
Qy	723	AlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMet	742
Db	2536	GCCATCAAGGTGTCTGCTGCTCACAGGCGCCGCTACTTCTGGGTAACCTTACCCGGAATG2595	
Qy	743	AsnArgProValIleAlaSerLysAsnAspLysLeuGluLeuGlnAspCysLeuGluTyr	762
Db	2596	AAACGGCCCGTGTGGCTTCAAGAAATGATAAGCTCGAGCTGCAGGACTGCCTGGAGTTC2655	
Qy	763	AsnArgIleAlaLysLeuLysLysValGlnThrIleThrThrLysSerAsnSerIleLys	782
Db	2656	AGTAGGACAGCAAGTTAAAGAAAGTCAGACAAATACCACCAAGTCGAACCTCCATCAGA2715	
Qy	783	GlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysGluAspValLeuTyrCys	802
Db	2716	CAGGCAAAACCCAGCTTTTCCCTGTAGTCATGAATGCAAGGACGACGTTTTGTGTGC2775	
Qy	803	ThrGluLeuGluArgIlePheGlyPheProValHisTyrThrAspValSerAsnMetGly	822
Db	2776	ACTGAGCTCGAAAGATCTTCGGCTTCCCTGCTCCTACACGACGCTGTCACATCGGCG2835	
Qy	823	ArgGlyAlaArgGlnLysLeuLeuGlyArgSerTyrSerValProValIleArgHisLeu	842
Db	2836	CGCGCGCCCTCAGAAAGCTGCTGGCAGGCTCTGGAGTGTACCGGTTCATCAGACACCTG2895	
Qy	843	PheAlaProLeuLysAspTyrPheAlaCysGlu	853
Db	2896	TTTGGCCCTTGAAGGACTACTTTGCTGTGAA2928	
RESULT	11		
LOCUS	AY078427	4006 bp mRNA linear ROD 01-JUL-2002	
DEFINITION	Mus musculus DNA cytosine methyltransferase 3b6 (Dnmt3b6) mRNA, complete cds; alternatively spliced.		
ACCESSION	AY078427		
VERSION	AY078427.1	GI:21655120	
KEYWORDS	house mouse.		
SOURCE	Mus musculus		
ORGANISM	Mus musculus		
REFERENCE	1 (bases 1 to 4006)		
AUTHORS	Chen, T., Ueda, Y., and Li, E.		
TITLE	Dnmt3a2 encoded by transcripts from an intronic promoter of Dnmt3a is expressed abundantly in ES cells and germ cells		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 4006)		
AUTHORS	Chen, T., Ueda, Y., and Li, E.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-FEB-2002) Cardiovascular Research Center, Massachusetts General Hospital, 149 13th Street, Charlestown, MA 02129, USA		
FEATURES	Location/Qualifiers		

source 1. .4006  
/organism="Mus musculus"  
/strain="129S4/SvJae"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/map="84.0 cM"  
1. .4006  
/gene="Dmnc3b6"  
269. .2659  
/gene="Dmnc3b6"  
/function="de novo methylation"  
/note="contains PHD and PWWP domains; alternatively spliced"  
/product="DNA cytosine methyltransferase 3b6"  
/protein\_id="AA185481.1"  
/db\_xref="GI:21655121"  
/translation="MKQDSRLHLEBEGAGSECIIVNGNFSQSSDTKDAPEPPLE  
ALTEPVCPTETGRSSRLSKREVSSLLNYQDMTGDGRDDEVDDGSDILMPK  
LIRETKDTRTRSSPAVTRSHSGSLSEORAPITRGRORHHVQEPVPAIR  
SRRRASSASSTSPASVDPMEEVTPKSTSPVLSQDGOEGMDTQVASED  
GDSREYODKEFGIDLVWGKIKGFSWMPAMVSWKATSKROMPGRWQWFGDGF  
SEISADKLVALIG.FSOHNLATPNKLVSXKAMVHLEKARPAAGKTSSPESLED  
QKPLMEHMGKFRPTGIBGLKPKKQPPVNVKSKVRSDSNLEPRRENKSRRTTN  
DSASBSPPRRLKTNVSGKDRGEDESESRMASEVTNKNLEDRCLSCGKNPV  
FHLPRGGLCOSCRDRELLELFWYDEGYOSVTCVCEGRELILCSITSCRCFV  
LEPLVAGTADKADLOEPASCYMLPQRCGVLRRRDMRLQDPTTDPDEEFP  
KLYPAIPAKRRPIRYLSLFDGIATGYLVLEKELIKVEKYAISVCAESIIVGTWV  
EGQIKVNDYRKITTKNIEMWGPDLVIGSPENDLSNVNPAKGLYEGVGR.FPPEY  
HLNNTYRKGDNRPFPMFENYAMKVNKKDISRLACNPWIDIKVSAHRAFY  
FMGNLPRMNIIFEPFAYITDVSNMGRGARKLGRSWSEVIRHLEPLADPYACE"  
BASE COUNT 1001 a 1017 c 1075 g 913 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 7.69e-173 Length: 4006  
Score: 3415.50 Matches: 660  
Percent Similarity: 81.52% Conservative: 50  
Best Local Similarity: 75.77% Mismatches: 68  
Query Match: 74.80% Indels: 93  
DB: 10 Gaps: 11  
US-09-720-086-8 (1-853) x AY078427 (1-4006)  
QY 1 Metlysglyasprtharghisleuasnglygluiauaplaaglyayrgluaspsr 20  
269 ATGAAGGAGACAGCACATCTGATGAAGAAGAGGTGCCAGCGGTATGAGGAGGCC 328  
QY 21 Ileuvalaenglyalacyseraspglnseraspr-----serpro 35  
Db 329 ATTATCGTTATGGGAAGCTTCAAGTCCACAGACAGAGAGATGCTCCCTCACCC 388  
QY 36 Proilleuenglualle-----argthrprogluilearglyayrglu 50  
Db 389 CCACTCTTGAGGCAATCTGCACAGGCCAGTCTGCACACAGACAGACGCGCGCAGG 448  
QY 51 SerSerSerargleuserlyarigluvalserSerleuserlyThrGluaspleu 70  
Db 449 TCAAGCTCCCGCGCTGTCTTAAGAGGAGGTCTCCAGCTTCTGAATTAACACAGACATG 508  
QY 71 ThrGlyaspglyasprly-----gluaspglyasprlyserasprthrproval 86  
Db 509 ACAGAGATGAGACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 565  
QY 87 Metprolyaleuphearglu-----thrargthraargsergluserproalalal 103  
Db 566 ATGCCAAAGCTCACCCGAGACCAAGACACAGGAGCGCTCTGAAAAGCCCGCGCTGTC 625  
QY 104 ArgthraargleuasprSerValserSerarglyuarghisarproserproargser 123  
Db 626 CGAACCCGACATAGATAGGAGCTCCAGCTTGAGAGGCAAAAGAGCTCCCGCCGAAATC 685  
QY 124 Thrargglyayrglnglyayrgasnhisvalaspgluserprovalglupheproalathr 143

Db 686 ACCGAGGTGGAGGAGGCCACCATGTGACAGAGATACCTGTGAGTTCCGGCTAC 745  
QY 144 ArgSerleuarargargalathraAlaSerAlaGlythrProTrpProSerProSer 163  
Db 746 AGGTCTCGAGAGCTCGAGCATGCTCTTCAGCAAGCAGCCATGATATCCCTGCCACG 805  
QY 164 Serlyrleuthrileaspleuthraspsprthrgluasprthrisgllythrproglinser 183  
Db 806 -----GTGACCTTCATGMAAGAGTG-----ACACCTAAGAGC 838  
QY 184 SerSerThrProTyralaargleualagluaspserserlnglnglylmetgluserpro 203  
Db 839 GTGACGATCCCATGCTGACTTGACGACAGATGAGATCAGAGGATGATGATGATGATGAT 898  
QY 204 Gluvalgluilaasprserglyasprglyaspserserlnglnglylmetgluserpro 223  
Db 899 CAGGTGATGACAGACAGACAGATGAGACAGACAGATGATGATGATGATGATGATGATGAT 958  
QY 224 GlylleuglyaspleuvaltyrplylyleleuglypasesertrpProAlaMetVal 243  
Db 959 GGAATAGTACCTCGTGTGGGAAATCATCAGGCTTCTCTGTGCTGCTGCTGCTGCTGCTG 1018  
QY 244 ValsertrpysalathrserlysarigluAlaMetserglmetargtrpvalglnttrp 263  
Db 1019 GTGCTCGGAAAGCCACCTCCAGCGACAGCGCATGCCCGGAAATGCCCTGGGTACAGTGG 1078  
QY 264 PheGlyasprglyyspasesergluvalserAlaasprlyseuvalaleuuglyleuphe 283  
Db 1079 TTGGTATAGGCAAGTTTCTGAGATCTGTGACAACTGGTGGTCTGGGGCTGTCTC 1138  
QY 284 SerGlnhisPheasleuualathrPheasnlyseuvalserTyralaGlysalameTyr 303  
Db 1139 AGCCAGACCTTATATCGGTACTTCAATTAAGCTGTTCTTATAGAGGCAATGATAC 1198  
QY 304 HisAlaLeugluysalaargvalargaleglylythrPheProSerSerProGlyasp 323  
Db 1199 CACACTCTGAGAAAGCAGAGGTTCGAGTGGCAGAACCTTCTCCACAGCTCCGAGAG 1258  
QY 324 SerleuugluasprngluuyspPromeTleugluTrpAlaHisgllygllypeliysprothr 343  
Db 1259 TCACTGAGAGCAAGCTGAAAGCCCATCTGAGTGGGCCCAAGCGGTGCTTCAACCTACT 1318  
QY 344 GlylleuglyleuyspPromeasnThrGlnProvalAlaasnlyseuvalaarg 363  
Db 1319 GGGATCGAGGGCTCAACCCACAAAGACCAACCAATGTTAATAGTCAGAGTGGCT 1378  
QY 364 ArgAlaGlySerArglyleugluysararglytyrGluasnlystrthraargargthr 383  
Db 1379 CGTTCAGACAGTAGAAGCTTAGAACCCAGAGACGCGAAGAACAAAGTCGAAGACGACA 1438  
QY 384 AlaaspserseralathrserasprtyrcysproAlaProlysagleuysrthrasnlys 403  
Db 1439 ACCAATGACTCTGTGCTTGTAG--TCGCCCCACCAAGCGCTCAAGACAAATAGC 1495  
QY 404 TyraasnenglyysaspararglyaspgluaspglnserarggluglnmetAlaSerasp 423  
Db 1496 TAT--GCGGGAAGACCGAGGGAGAGATGAGAGCCGAAGACGATGGCTTCTGAA 1552  
QY 424 ValAlaasnlyserSerleuugluaspglyCysleuserCysglayrglyasapro 443  
Db 1553 GTCAACCAACAAAGGCAATCTGGAAGACCGCTGTTGCTCTGGAAGAAACACCT 1612  
QY 444 ValserPhehisproleuphegluglylyleucysglnthrCysargasparargPheleu 463  
Db 1613 GTGCTCTTCAACCCCTCTTGAAGGTGGCTGTGAGAGTGCCTCGGAGTGGCTTCTTA 1672  
QY 464 GluLeupheTyMetTyraasparaspglyTyrglnserTyrcysThrValCysCysglu 483  
Db 1673 GAGCTCTTTCATATGATGAGACGCTATAGTCTTACCTGACACGCTGTGTGAG 1732  
QY 484 GlyarggluleuLeuLeuysSerasnThrSerCysargCyspheCysvalgluys 503

```
Db 1733 GCGCGTGAACGCTGCTGTGCAGTAACACAAAGCTGCTGCAGATGCTTCTGTGTGGAGTGT 1792
Qy LeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaValLeuGluGlnProTrpSer 523
Db 1793 CTGGAGGTGCTGTGGCGCAGGCACACAGCTGAGGATGCCAAGCTGCAGGAACCTCTGGAGC 1952
Qy CysTrpMetCysLeuProGlnArgCysHisGlyValLeuArgArgGlyAspTrpAsn 543
Db 1953 TCGTATATGCTCTCCCTCAGCGCTGCTGAGGTCTCTCCGACGAGGAAAGATTGAAC 1912
Qy ValArgLeuGlnAlaPhePheThrSerAspThrGlyLeu---GluTrpGluAlaProLys 562
Db 1913 ATGGCGCTGCAAGACTTCTTCACTACTGATCTCTGACCTGGAAGATTTGAGCCCAAG 1972
Qy LeuTrpProAlaIleProAlaAlaArgArgProIleArgValLeuSerLeuPheAsp 582
Db 1973 TTGTATCCAGCAATCTCTGACGCCAAAGAGGAGCCATTTAGAGTCTCTGTCTGTGAT 2032
Qy GlyIleAlaThrGlyTrpLeuValLeuLysGluLeuGlyIleLysValGlyLysTrpVal 602
Db 2033 GGAATTGCAACGGGTACTTGTGTCTCAAGGAGTTGGGTATTAAAGTGGAAAGTACATT 2092
Qy AlaserGluValCysGluGluSerIleAlaValGlyThrValLysHisGluGluAsnIle 622
Db 2093 GCCTCCGAGTCTGTGCAGATCTCATCGCTGTGGAACTGTTAAGCATGAAGGCCAGATC 2152
Qy LysTrpValAsnAspValArgAsnIleThrLysLysAsnIleGluGluTrpGlyProPhe 642
Db 2153 AATATGTCAATGATCTCGGAAATCCCAAGAAATATATGAAGATGGGGCCCGTTTC 2212
Qy AspLeuValIleGlySerProCysAsnAspLeuSerAsnValAsnProAlaArgLys 662
Db 2213 GACTTGTGATTTGGTGAAGCCCATGCAATGATCTCTTAAGCATCAATCTTCCCGCCGAAA 2272
Qy GlyLeuTrpGluGlyThrGlyArgLeuPheGluPheTrpHisLeuAsnTrpSer 682
Db 2273 GGTATTATGAGGCGACAGGAGGCTCTTCTTCGAGTTTACCACTTCTGCAATATATACC 2332
Qy ArgProLysGluGlyAspAspArgProPheTrpMetPheGluAsnValAlaIleMet 702
Db 2333 CGCCCCAAGGAGGCGCAACCGTCCATCTTCTTGTGATGTTGAGAAATGTTGTGGCCATG 2392
Qy LysValGlyAspLysArgAspIleSerArgPheLeuGluCysAsnProValMetIleAsp 722
Db 2393 AAAGTGAATGACAGAAGAAGACATCTCAAGATCTCTGGCATGTAAACCCAGTGCATCGAT 2452
Qy AlaIleValSerAlaAlaHisArgAlaArgTrpPheTrpGlyAsnLeuProGlyMet 742
Db 2453 GCCATCAAGGTGCTGTGCTGCTCACAGGCCCGGTACTTCTGGGGTAACTACCCGGATG 2512
Qy AsnArgProValIleAlaSerLysAsnAspLysLeuGluLeuGlnAspCysLeuGluTrp 762
Db 2513 AAC----- 2515
Qy AsnArgIleAlaLysLeuLysValGlnThrIleThrThrLysSerAsnSerIleLys 782
Db 2515 ----- 2515
Qy GlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysGluAspValLeuTrpCys 802
Db 2515 ----- 2515
Qy ThrGluLeuGluArgIlePheGlyPheProValHisTrpThrAspValSerAsnMetGly 822
Db 2516 -----AGGATCTTTCGGTCTTCCCTGCTCACTACACGAGCGTGTCAACATGGGC 2563
Qy ArgGlyValArgGlnLysLeuGluValArgSerTrpSerValProValIleArgHisLeu 842
Db 2564 CGCGGCGCCGTCAGAACTGCTGGCAGGTCCTGGAGTGTACCGGTCTATCAGACACTG 2623
Qy PheAlaProLeuLysAspTrpPheAlaCysGlu 853
Db 2624 TTGGCCCTTGAAGGACTACTTTCCTCTGTGAA 2656
```

## RESULT 12

AF151975  
LOCUS AF151975 4034 bp mRNA linear ROD 08-JUN-2000  
DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 7

ACCESSION AF151975  
VERSION AF151975.1 GI:8347134

KEYWORDS Mus musculus  
SOURCE Mus musculus

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 4034)  
AUTHORS Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and Shen, Y.

TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing isoforms in mouse embryonic tissue

JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 4034)  
Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and Shen, Y.

TITLE Direct Submission  
JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular Biology, Institution of Basic Medical Sciences, Chinese Academy of Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005, P.R. China

FEATURES  
Location/Qualifiers

1..4034  
/organism="Mus musculus"  
/strain="KM"  
/db\_xref="taxon:10090"  
/dev\_stage="8-9 day old embryo"

gene  
1..4034  
/gene="Dnmt3b"

CDS  
297..2687  
/gene="Dnmt3b"  
/function="methylates cytosine in DNA"  
/notes="alternatively spliced"  
/codon\_start=1  
/product="DNA cytosine-specific methyltransferase isoform 7"

protein\_id="AAF74521.1"  
/db\_xref="GI:8347135"  
translation="MKGDSRLHNEEGAGYECIIVNGFNFSDDSSDTKDAPSPPV  
AICTEPVCTPTGRSSRLSKREYSSLLANTQDMTGDDGDDVDGNGSDILAK  
LTERKDTSTSESPAPVTPHNSGTSLEQRASPRITRGROGRHHVQVYVPEPAT  
SRERRASSASTPSSPASVDPMENTPKSVTPSDVLSQDQEGMDTTOVAESLY  
GDSTEQDKKEFGIGLVKIGFSWAPAMVSKAKSKQAMFCMRWVQFGDGP  
SEISAKLVALGLFSQHPNLATFNKLVSKAMHTLEKARVAGTKTSSFGSLSD  
QLKPMLEWAGGFAFTGTEGLKFNKQPVVNGSKVRSDNRNLEPRRENKRRTTN  
DSAESPPPKALKINTSYGDKRGDEDESRERMASEVITNNKGLNLEDCSCCKKNPV  
FHLPEGGUCQSRDRFLELFYDDEYQSYCTVCCGRELILCSNCCRCFCVSC  
LEVLVAGTAEDAKLQEPWSCYKLPORCHGVLRKKDNNMLQDFFTPDLEEFEP  
PKLYPAIPAAKKRPIRVLSLDGATGVLVLELGIKVEKYITASEVCAESITAVTVEH  
EGQIKYVNDVRKTKKNIIEGPPDLVIGGSPCNLSNVNPKRGKGLYEGTGLFEY  
HCLNTPRKEGNDPFPFMMFENVVAMKNDKKDISRLACNPVMDAIKVSAAHARY  
FWGNLFGNRIFGFPAHYTDVSNMGRGAKQLGRSWSVPVIRHLFAPLKYDFACE"

BASE COUNT 991 a 1042 c 1097 g 904 t

ORIGIN  
Alignment Scores:  
Pred. No.: 2,98e-172 Length: 4034  
Score: 3404.50 Matches: 659  
Percent Similarity: 81.40% Conservative: 50  
Best Local Similarity: 75.66% Mismatches: 69  
Query Match: 74.56% Indels: 93  
DB: 10 Gaps: 11

US-09-720-086-8 (1-853) x AF151975 (1-4034)

Qy 1 MetLysGlyAspThrArgHisLeuAsnGlyGluAspAlaGlyGlyArgGluAspSer 20

Db	297	ATGAAGGAGACACAGACATCTGTAATGAAGAAAGACGGTCCAGAGGGATATGAGAGATGC	356
Oy	21	IIleuValAsnGlyAlaCysSerAspGlnSerSerAsp-----SerPro	35
Db	357	ATTATTCGTTAATGGAACTTCAGTGACCAAGTCCCTCAGACAGAAAGATGCTCCCTCACCC	416
Oy	36	ProIleuGlnAlaIle-----ArgThrProGlnIleArgGlyAArgArg	50
Db	417	CCAGCTTGAGGAGCAATCTGCACAGACCAGTCTGCACACCAAGAACAGAGCCGCGAGG	476
Oy	51	SerSerSerArgLeuSerIleArgGluValSerSerLeuLeuSerTyrThrGlnAspLeu	70
Db	477	TCAAGCTCCGGCGCTGTTAAGAGGAGAGTCTCCAGCTTCTGAATTACAGAGACAGCATG	536
Oy	71	ThrGlyAspGlyAspGly-----GluAspGlyAspGlySerAspThrProVal	86
Db	537	ACAGAGATGAGACACAGATGATGAAAGTGAATGAAATGAGAAATGGCTCTGATATT---CTA	593
Oy	87	MetProIysLeuPheAspGlu-----ThrArgThrArgSerGlnSerProAlaVal	103
Db	594	ATGCCAAAGCTCACCCGTGAGACCCAGACACACAGACACCGCTCTGAAAGCCCGGCTGTC	653
Oy	104	ArgThrArgAsnAsnAsnSerValSerSerArgGluArgHisArgProSerProArgSer	123
Db	654	CGAACCCGACATACCAATGGGACCTCCAGCTTGGAGAGCAAGAAAGCTCCCCCGAATC	713
Oy	124	ThrArgGlyArgGlnGlyArgAsnHisValAspGluSerProValGluPheProAlaThr	143
Db	714	ACCGAGGTGGCGAGGGCCGCAACATGTGCAGAGATACCGTGTGAAGTTTCCGGCTAC	773
Oy	144	ArgSerLeuArgArgValAlaThrAlaSerAlaGlyThrProTyrProSerProProSer	163
Db	774	AGGCTCGGACACTCGAGCATGCTCTTCGCAAGACCCCATGTCTATCCCTCCCGCAGC	833
Oy	164	SerTyrLeuThrIleAspLeuThrAspAspThrGluAspThrHisGlyThrProGlnSer	183
Db	834	-----GTCCACTTCATGAGAAAGTG-----ACACTTAAGAGC	866
Oy	184	SerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMetGluSerPro	203
Db	867	GTCAGTACCCATCAGTTGACTTGAGCCAGACAGATGAGAGATCAGGAGGTATGATGATCACA	926
Oy	204	GlnValGluAlaAspSerGlyAspGlyAspSerSerGluTyrGlnAspGlyValGluPhe	223
Db	927	CAGGTGATGACGAGACCATATATGGAGACAGACAGATCAGGATCATATAAGATTT	986
Oy	224	GlyIleGlyAspLeuValITrPGlyLysIleLysGlyPheSerTyrTrpProAlaMetVal	243
Db	967	GAATAGGTGACCTCGTGTTGGGAAATCAAGGGCTTCTCTGGTGGCTGCACATGGTG	1046
Oy	244	ValSerTrpLysAlaThrSerLysArgGlnAlaMetSerGlyMetArgTyrValGlnTrp	263
Db	1047	GTGTCTCTGGAAAGCACCTCCAGACGACAGGCCATGCCCCGAATGGCTGGGTACAGCG	1106
Oy	264	PheGlyAspGlyLysPheSerGluValSerIleAspLysLeuValAlaLeuGlyLeuPhe	283
Db	1107	TTTGGTATGCGCAAGTTTCTGAGATCTCTGTCGACAAACTGGTGGCTCTGGGCGTGTTC	1166
Oy	284	SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgGlyValaMetTyr	303
Db	1167	AGCCAGCATTTAATCTGGCTACCTTCATTAATACCTGGTTCTTATGAAAGCCATGTAC	1226
Oy	304	HisAlaLeuGlnLysAlaArgValArgAlaGlyLysThrPheProSerSerProGlyAsp	323
Db	1227	CACACTCTGGAGAAAGCCAGGGTTCCAGCTGGCAAGACCTTCTCCAGCAGTCTTGAAGAG	1286
Oy	324	SerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGlyPheLysProThr	343
Db	1287	TCACTGAGAGACACAGCTGAAGCCATGCTGGAAGTGGCCACGATGGCTTCAAGCTACT	1346
Oy	344	GlyIleGluGlyLeuLysProAsnAsnThrGlnProValValAsnLysSerLysValArg	363
Db	1347	GGGATCGAGGGCTCTCAACCCCAAGAAACCAACAGTGGTTAATTAAGTCGAAGGTGGT	1406

QY	364	ArgAlaGlySerThrArgIysLeuGluSerArgIysIleuValGluValThrArgArgArgThr	383
Db	1407	CGTTCAGACAGTGAAGAACTTAGAACCCAGAGAACCGCGAAGAAAGATCGAAGACGCACA	1466
QY	384	AlaAspAspSerLeuAlaThrSerAspTyrCysProAlaProLysArgLeuLysThrAsnCys	403
Db	1467	ACCAATGACTCTGCTGCTTCTGAG---TCCCCCACCACCCAGCGCTTCAACACAAATAGC	1523
QY	404	TyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluGlnMetAlaSerAsp	423
Db	1524	TAT---GCGCGAAGAGACCGAGGGAGGATGACGAGAGCCGAGAACCGATGGCTTCGAA	1580
QY	424	ValAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerCysGlyArgLysAsnPro	443
Db	1581	GTCACCAACACAAAGGCGACTCGAAGACCGCTGTGTCTCTGGAAAGAAAGAACCTT	1640
QY	444	ValSerPheHisProLeuPheGluGlyLysLeuCysGlnThrCysArgAspArgPheLeu	463
Db	1641	GTGTCTCTCCACCCCTCTTTGAGGGTGGGCTCTGTCAAGTGTCCGGAGATCGCTTCTTA	1700
QY	464	GluLeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThrValCysCysGlu	483
Db	1701	GAGCTCTTTACAAAGTATGATGATGAGAGAGCGCTATCACTCTACTGACCGCTGGCTGTAG	1760
QY	484	GlyArgGluLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCysValGluCys	503
Db	1761	GCGCGTAGAAGCTGCTGCTGTACAGTAACAAACCTGTCAGATCTTGTGTGGAGTGT	1820
QY	504	LeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnGluProThrSer	523
Db	1821	CTGAGAGGCTGGTGGGCGGACGACGACGCTGAGAGTGCACCACTGCAGAACCTCGAGC	1880
QY	524	CysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgArgLysAspTyrAsn	543
Db	1881	TGCTATATGTGCTCCTCCTCAGCGCTCCATGGGGTCTCTCCAGCGAGAAAGATTGGAAC	1940
QY	544	ValArgLeuGlnAlaPhePheThrSerAspThrGlyLeu--GluTyrGluAlaProLys	562
Db	1941	ATGCGCGTGCAGACACTTCTTCACTACTGATCTCCAGACTCGAAGAAATTGAGCCACCCAG	2000
QY	563	LeuTyrProAlaIleProAlaAlaArgArgArgProIleArgValLeuSerLeuPheAsp	582
Db	2001	TTTGACCCAGCAATTCCTGAGGCAAAAGAGGCCCATTAAGTCTCTCTCTGTTGAT	2060
QY	583	GlyIleAlaThrGlyTyrIleLeuValLeuLysGluLeuGlyIleLysValGlyLysTyrVal	602
Db	2061	GGAATTGCACAGGGGTACTTGGTGTCTCAAGAGATTGGATTAAAGTGGAAAAGTACTT	2120
QY	603	AlaSerGluValCysGluGluSerIleAlaValGlyThrValLysHisGluGlyAsnIle	622
Db	2121	GCCTCCGAAGCTGTGAGAGTCCATCGCTGTGGGAAGCTTTAAACATGMAAGGCCAGATC	2180
QY	623	LysTyrValAsnAspValAlaGlnIleThrLysLysAsnIleGluGluThrGlyProPhe	642
Db	2181	AAATATGTCATGACGTCGCCGAAAAACACCAAGAAAAAATATTGAAGATGGGGCCGCTTC	2240
QY	643	AspLeuValIleGlyLysSerProCysAsnAspLeuSerAsnValAsnProAlaArgLys	662
Db	2241	GACTTGGGATTTGGTGGAAACCCATCAATGATCTCTCAAGCTCAATCTCCGCCGANA	2300
QY	663	GlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPheTyrHisLysLeuAsnTyrSer	682
Db	2301	GGTTTATATGAGGGCACAAGAAAGGCTCTTCTTGAGTTTATACACTTGTGATATTATAC	2360
QY	683	ArgProLysGluGlyAspAspArgProPheThrPheCysPheGluAsnValValAlaMet	702
Db	2361	CGCCCCAAGAGGGCGACACACGCTCATCTTCTGGATGTTGCAAAAGTTGTGGCCATG	2420
QY	703	LysValGlyAspLysArgAspIleSerArgPheLeuGluCysAsnProValMetIleAsp	722
Db	2421	AAAGGAATGACAAAGAAAGCATCTCAAGATTCCTGGCATGTAAACCCAGTAGATGATCAT	2480

Qy 723 AlaIleValSerAlaHisArgAlaArgTy-PheTrpGlyAsnLeuProGlyMet 742  
Db 2481 GCATCAGGTGCTGCTCTACAGGCGCGTACTCTGGGGTAACCTACCGGATG 2540  
Qy 743 AsnArgProValIleAlaSerLysAsnAspLysLeuGlnAspCysLeuGluTyr 762  
Db 2541 AAC----- 2543  
Qy 763 AsnArgIleAlaLysLeuLysLysValGlnThrIleThrLysSerAsnSerIleLys 782  
Db 2543 ----- 2543  
Qy 783 GlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysGluAspValLeuTrpCys 802  
Db 2543 ----- 2543  
Qy 803 ThrGluLeuGluArgIlePheGlyPheProValHisTyrThrAspValSerAsnMetGly 822  
Db 2544 -----AGGATCTTCGGCTTCCTGCTCACTACAGGAGCTGTCCAACATGGC 2591  
Qy 823 ArgGlyAlaArgGlnLysLeuGluLysLeuGluValProValIleArgHisLeu 842  
Db 2592 CGCGGCGCGCTCAGAGCTGCTGGCAGGTCTCTGGAGTGTACCGGTCTATCAGACACTG 2651  
Qy 843 PheAlaProLeuLysAspTy-PheAlaCysGlu 853  
Db 2652 TTTGCCCTTGAAGGACTACTTTGCCCTGTGAA 2684

RESULT 13  
AF151972 4149 bp mRNA linear ROD 08-JUN-2000  
LOCUS Mus musculus DNA cytosine-specific methyltransferase isoform 4  
DEFINITION (Dnmt3b) mRNA, complete cds.  
ACCESSION AF151972  
VERSION AF151972.1 GI:8347125  
KEYWORDS  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 4149)  
AUTHORS Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and Shen, Y.  
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing isoform in mouse embryonic tissue  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4149)  
AUTHORS Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and Shen, Y.  
TITLE Direct Submission  
JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular Biology, Institution of Basic Medical Sciences, Chinese Academy of Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005, P. R. China

FEATURES  
source  
1..4149  
/organism="Mus musculus"  
/strain="KM"  
/db\_xref="taxon:10090"  
/dev\_stage="8-9 day old embryo"  
1..4149  
/gene="Dnmt3b"  
412..2802  
/gene="Dnmt3b"  
/function="methylates cytosine in DNA"  
/note="alternatively spliced"  
/codon\_start=1  
/product="DNA cytosine-specific methyltransferase isoform 4"  
protein\_id="AAF74518.1"  
/db\_xref="GI:8347126"  
/translation="MKGDSRLHNEEGAGSYEECIIVNGNFSPQSSDTKDAPSPVLS  
AICTPVCITETGRSSRLSKREVSSLLNTQDMDGDDRDEYDDGNGSDILMPK

BASE COUNT 1013 a 1082 c 1123 g 931 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3 08e-172 Length: 4149  
Score: 3404.50 Matches: 659  
Percent Similarity: 81.40% Conservative: 50  
Best Local Similarity: 75.66% Mismatches: 69  
Query Match: 74.56% Indels: 93  
DB: 10 Gaps: 11  
US-09-720-086-8 (1-853) x AF151972 (1-4149)  
Qy 1 MetLysGlyAspThrArgHisLeuAsnGlyGluLeuAlaGlyArgGluAspSer 20  
Db 412 ATGAAGGGAGACACGACACATCTGAATGAAGAGGGTCCAGCGGTATGAGGAGTGC 471  
Qy 21 IleLeuValAsnGlyAlaCysSerAspGlnSerSerAsp-----SerPro 35  
Db 472 ATTATCGTTAATGGAACTTTCAGTGACCACTCTCAGACAGAGGATGTCCTCCATCCC 531  
Qy 36 ProfileLeuGluAlaIle-----ArgThrProGluIleArgGlyArgArg 50  
Db 532 CCAAGTCTTTGGAGGCAATCTGCACAGAGCCAGTCTGCACACAGACAGAGCCGAGG 591  
Qy 51 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuSerTyrThrGlnAspLeu 70  
Db 592 TCAAGCTCCCGCTGTCTAAGAGGAGGTCTCCAGCTTCTGAATACACGAGGACATG 651  
Qy 71 ThrGlyAspGlyAspGly-----GluAspGlyAspGlySerAspThrProVal 86  
Db 652 ACAGGAGATGAGACAGAGATGATGAAGTAGATGATGGAATGGCTCTCATATT---CTA 708  
Qy 87 MetProLysLeuPheArgGlu-----ThrArgThrArgSerGluSerProAlaVal 103  
Db 709 ATGCCAAGCTCACCCGTGAGACCAAGGACACCAAGGACCGCTCTGAAAGCCCGCTGTC 768  
Qy 104 ArgThrArgAsnAsnAsnSerValSerSerArgGluArgHisArgProSerProArgSer 123  
Db 769 CGAACCCACATAGCAATGGACCTCCAGCTTGGAGAGGCAAGAGCCTCCCCAGATC 828  
Qy 124 ThrArgGlyArgGlnGlyArgAsnHisValAspGluSerProValGluPheProAlaThr 143  
Db 829 ACCCGAGGTGGCAGGCGCCACCATGTCAGGAGTACCCCTGTGGAGTTTCGGGTACC 888  
Qy 144 ArgSerLeuArgArgArgAlaThrAlaSerAlaGlyThrProTrpProSerProProSer 163  
Db 889 AGGTCTCGGAGAGCTGAGCATCGTCTTCAGCAAGACACCGCATGGTGCATCCCCGTCAGC 948  
Qy 164 SerTyrLeuThrIleAspLeuThrAspAspThrGluAspThrHisGlyThrProGlnSer 183  
Db 949 -----GTCCACTTTCATGGAAGAAGTG-----ACACCTAAGAGC 981  
Qy 184 SerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyGlyMetGluSerPro 203  
Db 982 GTCAGTACCCCATCAGTTCACTTGAGCCAGGATGGAGATCAGGAGGGTATGGATACCACA 1041  
Qy 204 GlnValGluAlaAspSerGlyAspGlyAspSerSerGluTyrGlnAspGlyLysGluPhe 223  
Db 1042 CAGGTGGATGCAGAGAGCATATATGAGACACACAGATATCAGGATGATTAAGNGTTT 1101



QY 224 GlyIleGlyAspLeuValTyrGlyIleGlySerPheSerTyrProAlaMetVal 243  
 Db 1102 GGAATAGGTGACCTGGTGGGAAAGATCAAGGGCTTCTCGTGGCCCATGGTGTG 1161  
 QY 244 ValSerTyrPheValThrSerLeuArgGlnAlaMetSerGlyMetArgTyrValGlnTyr 263  
 Db 1162 GTGCTCTGGAAAGCACCTCCAGCGACAGGCCATGCCGGAATGCCGTGGGTACAGTGG 1221  
 QY 264 PheGlyAspGlyLysPheSerGluValSerAlaAspLysLeuValAlaLeuGlyLeuPhe 283  
 Db 1222 TTTGGTGTATGCAAGTTTCTTGAGATCTCTGCTGCAAACTGGTGGCTCTGGGGCTGTTC 1281  
 QY 284 SerGlnHisPheLeuLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyr 303  
 Db 1282 AGCAGACACTTAACTGGCTACCTTCAATAGCTGTCTTATAGCAAGCCATGTAC 1341  
 QY 304 HisAlaLeuGlnLysValArgValArgAlaGlyLysThrPheProSerSerProGlyAsp 323  
 Db 342 CACACTCTGGAGAAAGCCAGGGTTGAGCTGGCAAGACCTTCTCCAGCAGTCTGGAGAG 1401  
 QY 324 SerLeuGlnAspGlnLeuLysProMetLeuGlnTyrAlaHisGlyGlyPheLysProThr 343  
 Db 1402 TCACCTGGAGGACCAAGCTGAACCCATGCTGAGTGGGCCCAAGGTGCTTCAAGCTTACT 1461  
 QY 344 GlyIleGlyLysLeuLysProAsnAsnThrGlnProValValAsnLysSerLysValArg 363  
 Db 1462 GGGATCGAGGGCTCAAAACCCACAGAGAGCCAGTGTAAATAGTCCAGAGGTGCGT 1521  
 QY 364 ArgAlaGlySerArgLysLeuGlnSerArgLysTyrGlnAsnLysThrArgArgArgThr 383  
 Db 1522 CGTTCAGACATAGAGAACTTGAACCCAGGAGACCCGAGAAACAAAGTCGAAGAGCACA 1581  
 QY 384 AlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLeuLysThrAsnCys 403  
 Db 1582 ACCAATGACTCTGCTGCTTGTAG---TCCCCCACCACCAAGCGCTCAAGACAAATAGC 1638  
 QY 404 TyrAsnAsnGlyLysAspArgGlyAspGlnAspGlnSerArgGlnMetAlaSerAsp 423  
 Db 1639 TAT---GGCGGGAAGAGACCGAGGGAGATGAGAGACCCAGAAACGATGGCTTGTAA 1695  
 QY 424 ValAlaAsnAsnLysSerSerLeuGlnAspGlyCysLeuSerCysGlyArgLysAsnPro 443  
 Db 1696 GTCAACCAACAAAGGCGCAATCTGAGAACCGCTGTTGCTCTGGAAAGAAAGAACCT 1755  
 QY 444 ValSerPheHisProLeuPheGlnGlyGlyLeuCysGlnThrCysArgAspArgPheLeu 463  
 Db 1755 GTGCTCTTCCACCCCTCTTTGAGGGTGGGCTGTCTCAGAGTCCCGGGAATGCTTCTTA 1815  
 QY 464 GluLeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThrValCysCysGln 483  
 Db 1816 GAGCTCTTCTCATGTATGATGAGAGCGGCTATCAGTCTTACCTGACCGGTGTGTGAG 1875  
 QY 484 GlyArgGluLeuLeuLeuCysSerAsnThrSerCysArgCysPheCysValGlnCys 503  
 Db 1876 GGCGGTGAACCTGCTGTGTGACGTAAACAAAGCTGCTCAGATGCTTCTGTGTGAGAGT 1935  
 QY 504 LeuGlnValLeuValGlyThrGlyThrAlaAlaGlnAlaLysLeuGlnGluProThrSer 523  
 Db 1936 CTGGAGGTGCTGGTGGGCGACAGCCAGCTGAGATGCGACAGTCCAGAAACCTTGGAGC 1995  
 QY 524 CysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgArgLysAspTyrAsn 543  
 Db 1996 TGCATATGTGCTCTCCACAGGGTGCATGGGGTCTCCGACGCAAGAAAGATTGGAAAC 2055  
 QY 544 ValArgLeuGlnAlaPhePheThrSerAspThrGlyLeu---GluTyrGlnAlaProLys 562  
 Db 2056 ATGGCCCTGCAGAGACTTCTTCACTACATCTGACCTGCAAGAAATTGAGCCACCAAG 2115  
 QY 563 LeuTyrProAlaIleProAlaAlaArgArgArgProIleArgValLeuSerLeuPheAsp 582  
 Db 2116 TTGTACCAAGCAATCTCTGCGCCCAAGAGAGGCCCATTTAGATCTCTGTGTGTAT 2175  
 QY 583 GlyIleAlaThrGlyTyrLeuValLeuLysGluLeuGlyIleLysValGlyLysTyrVal 602

Db 2176 GGAATTCAACGGGGTACTGTGTCTGACAGAGTGGGTATTAAGTGAAGAAATGACATT 2235  
 QY 603 AlaSerGlnValCysGlnGlnSerIleAlaValGlyThrValLysHisGlnGlyAsnIle 622  
 Db 2236 GCCCTCCAAAGTCTGTGAGAGTCCATGCTGTGGAACTGTTAAGCATGAAGGCACATC 2295  
 QY 623 LysTyrValAsnAspValArgAsnIleThrLysAlaAsnIleGlnGlnTyrGlyProPhe 642  
 Db 2296 AAATATGTCAATGACGTCCGGAAATACACCAAGAAATATTGAAGATGGGGCCGCTTC 2355  
 QY 643 AspLeuValIleGlyGlySerProCysAsnAspLeuSerAsnValAsnProAlaArgLys 662  
 Db 2356 GACTTGTGTATGTGTGAAGCCCATGCAATGATCTCTAACCCTCAATCTCCCGGAAA 2415  
 QY 663 GlyLeuTyrGlnGlyThrGlyArgLeuPheGlnGluPheTyrHisLeuLeuAsnTyrSer 682  
 Db 2416 GGTTATATGAGGACACAGAAAGCTTCTTTCGAGTTTTCACCTTGTCAATTATAC 2475  
 QY 683 ArgProLysGlnGlyAspAspArgProPhePheTyrMetPheGlnAsnValAlaIleMet 702  
 Db 2476 CGCCCAAGAGGGGCGACAAACCGTCCATCTTGTGATGTTCCGAGATGTTGTGGCCATG 2535  
 QY 703 LysValGlyAspLysArgAspIleSerArgPheLeuGlnCysAsnProValMetIleAsp 722  
 Db 2536 AAAGTGAATGCAAGAAAGACATCTCAAGATTCCTGGCATGTAAACCAAGATGATGAT 2595  
 QY 723 AlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTyrProGlyAsnLeuProGlyMet 742  
 Db 2596 GCCATCAAGGTGTGTGCTGCTCACAGGGCCCGGTACTTCTGGGGTAACTTACCAGGAATG 2655  
 QY 743 AsnArgProValIleAlaSerLysAsnAspLysLeuGlnLeuGlnAspCysLeuGlnTyr 762  
 Db 2656 AAC----- 2658  
 QY 763 AsnArgIleAlaLysLeuLysValGlnThrIleThrLysSerAsnSerIleLys 782  
 Db 2658 ----- 2658  
 QY 783 GlnLysAsnGlnLeuPheProValValMetAsnGlyLysGluAspValLeuTyrCys 802  
 Db 2658 ----- 2658  
 QY 803 ThrGlnLeuGlnValArgIlePheGlyPheProValHisTyrThrAspValSerAsnMetGly 822  
 Db 2659 -----AGATCTTGGCTTCCCTGCTCCTCACACGACGCTGCCAATGGGC 2706  
 QY 823 ArgGlyAlaArgGlnLysLeuLeuGlyArgSerTyrSerValProValIleArgHisLeu 842  
 Db 2707 CGCGGCGCCGCTCAGAAAGCTGCTGGCAGAGTCTCGAGTGAACCGTCAATCAGACACTG 2766  
 QY 843 PheAlaProLeuLysAspTyrPheAlaCysGln 853  
 Db 2767 TTTGCCCTTGAAGACTACTTTCCTGTGAA 2799  
 RESULT 14  
 AF068628 3946 bp mRNA linear ROD 06-DEC-1999  
 LOCUS AF068628  
 DEFINITION Mus musculus DNA cytosine-5' methyltransferase 3B3 (Dnm3b) mRNA,  
 alternatively spliced, complete cds.  
 ACCESSION AF068628  
 VERSION AF068628.2 GI:6449473  
 KEYWORDS  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 3946)  
 AUTHORS Okano, M., Xie, S. and Li, F.  
 TITLE Cloning and characterization of a family of novel mammalian DNA  
 (cytosine-5' methyltransferases  
 Nat. Genet. 19 (3), 219-220 (1998)  
 JOURNAL MEDLINE 98324766

```

9662389
2 (bases 1 to 3946)
REFERENCE
AUTHORS Xie,S., Okano,M. and Li,E.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,
Charlestown, MA 02129, USA
3 (bases 1 to 3946)
REFERENCE
AUTHORS Okano,M., Chijiwa,T., Sasaki,H. and Li,E.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,
Charlestown, MA 02129, USA
REMARK
COMMENT Sequence update by submitter
FEATURES
On Nov 18, 1999 this sequence version replaced gi:3327983.
Location/Qualifiers
1..3946
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="2"
/map="84.0 cM"
1..3946
/gene="Dnmt3b"
1..3946
/misc_feature
1..3946
/gene="Dnmt3b"
/note="similar to EST sequences deposited in GenBank
Accession Numbers AA116694, AA119979, AA177277, AA210568,
AA407106, and AA575617"
269..2599
/gene="Dnmt3b"
/functions="de novo DNA methylation"
/note="alternatively spliced product; contains Cys-rich
region"
/codon_start=1
/product="DNA cytosine-5 methyltransferase 3B3"
/protein_id="AAC40180.2"
/db_xref="GI:6449474"
/translators="MGDSRLHNEEGASGYEECIIVNGNFSQSDTKDAPSPVLE
LITREKTPTRSPAPVTRHNSGTSLEORAPRITGRGRHVRVQIYVPEFPAIR
SRRRSSASTPWSPASVDFMEVTPKSVSTPSVLSQDQDQGMDDTQVDAESRD
GSDTQDDKDFGIGLVMGKIKGFWPMVAMVSWKATSKRQAMPQMRVQWQFGDGK
SEISAKLVALGFSQHNLAIFNKLVSRYKAMYHTLEKARVRAGTFFSSPSESLED
OLKPMLEWAGKPGPTGIEGLKPNKKQPKNSRRRTTNSAASESPPKRUKTNSYGG
KDGDESESRMASEVTNNKGLNEDCLSCGKNPVSPHPLFEGGLQSCDRFLFL
PYMDEGYSYCTVCEGRELLCNTSCRCFCVECLEVIVGAGTAEDAKLQEPMS
CYMCLPORCHVLRKDKMMLRLQDFTTDPDLSEEPFKLYPAIPAARRPIRVL
PDGIATLYLVKELGIVKEKYIASEVCAESIAGTVKHEGQIKYVNDVRKITKNIE
WGPFDLVIGSPCNDSNVNPAKGLYEGRIFFEFYHLLNRYRPEKGRNRPFWMF
ENVVMKRVNDKDISRFACNPVIMDAIKVSAHRARYFWGNLPGMNRIFGFPFAHYTD
VSNMGRARQKLGRSWSVPVIRHFLFLPKDYFACE"
BASE COUNT 983 a 1006 c 1056 g 901 t
ORIGIN

Alignment Scores:
Pred. No.: 2,22e-168 Length: 3946
Score: 3331.50 Matches: 646
Percent Similarity: 79.68% Conservative: 48
Best Local Similarity: 74.17% Mismatches: 64
Query Match: 72.96% Indels: 113
DB: 10 Gaps: 12

US-09-720-086-8 (1-853) x AF068628 (1-3946)
QY 1 MetLysGlyAspThrArgHisLeuAsnGlyGluGluAspAlaGlyGlyValGluAspSer 20
Db 269 ATGAAGGGAGACAGACAGACATCTGAATGAAGAAGAGAGGGTGCAGCGGGTATGAGGAGTGC 328
QY 21 IleLeuValAsnGlyAlaCysSerAspGlnSerSerAsp-----SerPro 35
Db 329 ATTATCGTTAATGGGAACCTTCAGTGACCACTCTCTCAGACACAGCAAGAGATGCTCCCTCACCC 388
QY 36 ProIleLeuGluAlaIle-----ArgThrProGluIleArgGlyValArg 50
Db 389 CCAGTCTTGAGGCANTCTGCACAGAGCCAGTCTGCACACACAGACACAGACGAGCGCGAGG 448

```

Qy 404 TyrAsnAngIyLysAspArgGlyAspGluAspGlnSerArgGluGlnMetAlaSerAsp 423  
 Db 1436 TAT---GGCGGGAGAGACCGAGGGAGATGAGAGAGACCGAGAAACGATGGCTTCTGAA 1492  
 Qy 424 ValAlaAsnAsnLysSerSerLeuGlnAspGlyCysLeuSerCysGlyArgLysAsnPro 443  
 Db 1493 GTCAACAAACAAGGGCAATCTGGAAGACCGCTGTTGCTCTGGGAAAGAAACCT 1552  
 Qy 444 ValSerPheHisProLeuPheGluGlyLysLeuCysGlnThrCysArgAspArgPheLeu 463  
 Db 1553 GTGCTCTTCCACCCCTCTTGGAGGGTGGCTGTCTGACAGTGGCCGGAGTCCCTTCTTA 1612  
 Qy 464 GluLeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThrValCysCysGlu 483  
 Db 1613 GAGCTCTTCTACATGATGATGATGAGACGGCTATCAGTCTCTAGTCAACCGTGTGCTGAG 1672  
 Qy 484 GlyArgGluLeuLeuLeuLeuCysSerAsnThrSerCysArgCysPheCysValGluCys 503  
 Db 4673 GGCGGTGAACCTGCTGTGTGACATCAACCAAGCTGCTGCAATGCTTCTGTGTGAGTGT 1732  
 Qy 504 LeuGluValLeuValGlyTyrGlyTyrAlaAlaGlnAlaLysLeuGlnGluProTrpSer 523  
 Db 1733 CTGAGAGGTGCTGGTGGGGAGGAGGACAGCTAGAGATGCCAAGCTGCAGAAACCTTGAGC 1792  
 Qy 524 CysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgArgLysAspTrpAsn 543  
 Db 1793 TGCTATATGCTCTCCCTCAGCGCTGCGATGGGGTCTCCAGCGCAGAAAGATTGGAAC 1852  
 Qy 544 ValArgLeuGlnAlaPhePheThrSerAspThrGlyLeu---GluTyrGluAlaProLys 562  
 Db 1853 ATGGCGCTGCAAGACTTCTTCACTATCTGACCTGACCGGAAGAAATTTGACCCACCAAG 1912  
 Qy 563 LeuTyrProAlaIleProAlaAlaArgArgArgProIleArgValLeuSerLeuPheAsp 582  
 Db 1913 TTGTAACCCAGCAATCTCTGCGAGCCCAAGAGGCCCATTAAGTCTCTCTCTTGTAT 1972  
 Qy 583 GlyIleAlaThrGlyTyrLeuValLeuLysGluLeuGlyIleLysValGlyLysTyrVal 602  
 Db 1973 GGAATTGCAACGGGGTACTTGTGTCTCAAGAGGTTGGTATTAAAGTGAAGAAAGACTT 2032  
 Qy 603 AlaSerGluValCysGluGlnSerIleAlaValGlyTyrValLysGluGlyAsnIle 622  
 Db 2033 GCCTCGCAAGCTGTGCAAGTCCATCTCTGTGGGACCTGTTAAGCATGAAGGCGCAGATC 2092  
 Qy 623 LysTyrValAsnAspValArgAsnIleThrLysLysAsnIleGluGlnTrpGlyProPhe 642  
 Db 2093 AAATATGCAATGACGTCCGAAATACCAAGAAATATTTGAAGATGGGGCCCTTC 2152  
 Qy 643 AspLeuValIleGlyGlySerProCysAsnAspLeuSerAsnValAsnProAlaArgLys 662  
 Db 2153 GACTTGGTGAATTTGGTGGAGGCCCATGATGATCTCTTAAGTCAATCTGCCCCGAA 2212  
 Qy 663 GlyLeuTyrGluGlyTyrGlyArgLeuPheGluPheTyrHisLeuLeuAsnTyrSer 682  
 Db 2213 GGTATATATGAGGGGACAGGAAGGCTCTTCTGAGTTTACCACTTGCTGATATATACC 2272  
 Qy 683 ArgProLysGluGlyAspAspArgProPhePheTrpMetPheGluAsnValValAlaMet 702  
 Db 2273 CGCCCAAGAGGGGAGAACCGTCCATCTTCTGAGATTTGGAAGATGTTGTGGCCATG 2332  
 Qy 703 LysValGlyAspLysArgAspIleSerArgPheLeuGluCysAspProValMetIleAsp 722  
 Db 2333 AAAGTGAATGACAGAAAGACATCTCAAGATTTCCGGAGCATGTAACCCAGTATGATCAT 2392  
 Qy 723 AlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMet 742  
 Db 2393 GCCATCAAGGTCTGCTGCTCAACGGGCGCGGTCTTCTGGGGTAACTACCCCGAATG 2452  
 Qy 743 AsnArgProValIleAlaSerLysAsnAspLysLeuGluAspCysLeuGluTyr 762  
 Db 2453 AAC----- 2455  
 Qy 763 AsnArgIleAlaLysLeuLysValGlnThrIleThrTyrSerAsnSerIleLys 782

Db 2455 ----- 2455  
 Qy 783 GlnGlyLysAngIlnLeuPheProValValMetAsnLysGluAspValLeuTrpCys 802  
 Db 2455 ----- 2455  
 Qy 803 ThrGluLeuGluArgIlePheGlyPheProValHisTyrThrAspValSerAsnMetGly 822  
 Db 2456 -----AGATCTTCGGCTTCCCTGCTCAGTACACAGGACGTCATCATGGGC 2503  
 Qy 823 ArgGlyAlaArgGlnLysLeuGlyLysArgSerTyrSerValProValIleArgHisLeu 842  
 Db 2504 CGCGGCCCGCTCAGAGCTGCTGGCAGGTCCTCGAGTGAACGGTCAATCAGACACTG 2563  
 Qy 843 PheAlaProLeuLysAspTyrPheAlaCysGlu 853  
 Db 2564 TTGCCCCCTTGAAGACTACTTGGCTGTGAA 2596  
 RESULT 15  
 AF151976 3974 bp mRNA linear ROD 08-JUN-2000  
 LOCUS AF151976  
 DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 8  
 ACCESSION AF151976  
 VERSION AF151976.1 GI:8347136  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 3974)  
 Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and  
 Shen, Y.  
 Cloning of full-length Dnmt3b cDNA and its alternative splicing  
 isoforms in mouse embryonic tissue  
 Unpublished  
 2 (bases 1 to 3974)  
 Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and  
 Shen, Y.  
 Direct Submission  
 Submitted (17-MAY-1999) Department of Biochemistry and Molecular  
 Biology, Institution of Basic Medical Sciences, Chinese Academy of  
 Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,  
 P. R. China  
 FEATURES  
 source  
 1. .3974  
 /organism="Mus musculus"  
 /strain="XM"  
 /db\_xref="taxon:10090"  
 /dev\_stage="8-9 day old embryo"  
 1. .3974  
 /gene="Dnmt3b"  
 297. .2627  
 /gene="Dnmt3b"  
 /function="methylates cytosine in DNA"  
 /note="alternatively spliced"  
 /codon\_start=1  
 /product="DNA cytosine-specific methyltransferase isoform 8"  
 /protein\_id="AAF74522.1"  
 /db\_xref="GI:8347137"  
 /translation="MKGDSRHINEREGASGEYECITVNGNFSQSSPTKDAAPRYE  
 ALCTEPCVTPETGRSSSRSLSKREYSLSLNTYODMTGDDRDDEVDNGSDILMK  
 LTRETDQTRRSRSPAVRTKHSNGTSSLSERASPRITRGQRHHVOEYVPEPATR  
 SRRRRSSASSTPWSPPASVDLFWMEVTPKSVSPVDLSGDGQEGADITQVASELY  
 GDSLEYODDDEFGIDLVWGKIGFPMWPMVVSMTATSKRQAMPGRWQMGDKF  
 SEISADKLVALVGFQHPNLATFNKLVSRYKAMVYTLERKAVAGAKTFSSPGESED  
 QIKPMLEMAHGPKPRTGIGIKPKKQPKNSRRRTTNDSAASESPPKRLKNSYGG  
 KDKGEDESRERMASEVTNKKGNLIERCLSCGKNVSPHPLREGGICQSCRPLFL  
 FMYWDEGYOSTVTCCEGRELDCNLTSCRCFVCELVLAGAGAEPAKIQEPPS  
 CYMCLPQRCHGVLRKRNMRLODFETDPLDEEPRLYLPAIPAKKRPRLVLSL  
 FDCIANGYLVLEKELGKVEKYIASVECAESIAGTVNHEGQIKYVNDVRKIKTKNIEE



Db 2121 AAATRTGCAATGACGTCGGAAAAATCAAGAAAAATTTGAAGATGGGGCCCTTC 2180  
 QY 643 AspleuVal11EG1yGlySerProCysAsnAspleuSerAsnValAsnProAlaArgLys 662  
 Db 2181 GACTTGATGATTGGTGGAGAGCCCATGATGATCTCTTAAGTCAATCCTGCCGCAA 2240  
 QY 663 GlyLeuTYRGLUGLYThrGLYArgLeuPheGluPheTYRHisLeuLeuAsnTYRSer 682  
 Db 2241 GGTATTATGAGGGGACAGGAAGCTCTTCGAGTTTACCACTTGCTGAATTATACC 2300  
 QY 683 ArgProLysGluGlyAspAspArgProPhePheTYRMetPheGluAsnValValAlaMet 702  
 Db 2301 CGCCCAAGAGAGCGGACACCGTCCATCTCTGAGATTGAGAAATGTTGGCCATG 2360  
 QY 703 LysValGlyAspLysArgAspLysSerArgPheLeuGluCysAsnProValMetLysAsp 722  
 Db 2361 AAAGGATGACAGAAAGACATCTCAAGATTCTCGGATGAAACCACTGATGATCAT 2420  
 QY 723 Ala11LeuValSerAlaAlaHisArgAlaArgTYRPhetTPGlyAsnLeuProGlyMet 742  
 Db 2421 GCCATCAAGGTGTCTGCTCCTACAGGCGCGGTACTTCTGGGTAACTTACCCGGAATG 2480  
 QY 743 AsnArgProVal11Leu1SerLysAsnAspLysLeuGluLeuGluAspCysLeuGluTYR 762  
 Db 2481 AAC----- 2483  
 QY 763 AsnArg11eAlaLysLeuLysLeuValGlnThr11eThrLysSerAsnSer11eLys 782  
 Db 2483 ----- 2483  
 QY 783 GlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysGluAspValLeuTYR 802  
 Db 2483 ----- 2483  
 QY 803 ThrGluLeuGluArg11ePheGlyPheProValHisTYRThrAspValSerAsnMetGly 822  
 Db 2484 -----AGGATCTTCGGCTTCCTGCTCACTACACGAGCTGTCCAAATGGGC 2531  
 QY 823 ArgGlyAlaArgGlnLysLeuLeuGlyArgSerTrpSerValProVal11eArgHisLeu 842  
 Db 2532 CGCGGCGCGCGCTGAGAACTGCTGGGCAAGTCTTGAGAGTGTACCGGTCACTACACACCTG 2591  
 QY 843 PheAlaProLeuLysAspTYRPhetAlaCysGln 853  
 Db 2592 TTGGCCCTTGAAGACTACTTGTCTGTGAA 2624

Search completed: November 22, 2002, 10:59:08  
 Job time : 4955.31 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 22, 2002, 03:58:32 ; Search time 5196.49 Seconds  
(without alignments)  
5107.633 Million cell updates/sec

Title: US-09-720-086-7  
Perfect score: 4935  
Sequence: 1 MPMSSSGPDTSSSAARE.....WSVPVIRHFLAPIKEYFACY 912

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 1451402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xid  
-O=/c/m2.1/USPTO.spool/US09720086/runat.18112002.092832.22195/app.query.fasta\_1.4252  
-DB=GenEmbl -QPM=fastap -SUFFIX=rge -MINMATCH=0 -LOOFCU=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45  
-DOCFALGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09720086.CCN 1.11308 @runat.18112002.092832.22195 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEJURY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THEADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-Ygapop=10 -Ygapext=0.5 -DELop=6 -DELext=7

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_hg: \*  
3: gb\_in: \*  
4: gb\_cm: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*

29: em\_vi: \*  
30: em\_hg\_hum: \*  
31: em\_hg\_inv: \*  
32: em\_hg\_other: \*  
33: em\_hg\_mus: \*  
34: em\_hg\_pin: \*  
35: em\_hg\_rod: \*  
36: em\_hg\_mem: \*  
37: em\_hg\_vrl: \*  
38: em\_sy: \*  
39: em\_higo\_hum: \*  
40: em\_higo\_mus: \*  
41: em\_higo\_other: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4935	100.0	3005	9 AF067972	AF067972 Homo sapi
2	4935	100.0	4258	9 AF331856	AF331856 Homo sapi
3	4740.5	96.1	4192	10 AF068625	AF068625 Mus muscu
4	4737.5	96.0	4094	10 BC007466	BC007466 Mus muscu
5	2541.5	51.5	123936	2 AC120824	AC120824 Rattus no
6	2413	48.9	2191	9 AK025230	AK025230 Homo sapi
7	2228.5	45.2	4163	10 AF151974	AF151974 Mus muscu
8	2227.5	45.1	168651	9 AC009474	AC009474 Homo sapi
9	2226	45.1	4135	10 AF068627	AF068627 Mus muscu
10	2225	45.1	4278	10 AF151970	AF151970 Mus muscu
11	2223.5	45.1	4223	10 AF151973	AF151973 Mus muscu
12	2221	45.0	4195	10 AF068626	AF068626 Mus muscu
13	2220	45.0	4338	10 AF151969	AF151969 Mus muscu
14	2211	44.8	4335	9 AF331857	AF331857 Homo sapi
15	2203	44.6	4267	9 AF176228	AF176228 Homo sapi
16	2199.5	44.6	176697	2 AC112586	AC112586 Rattus no
17	2196	39.5	4145	9 AF156488	AF156488 Homo sapi
18	1956	39.6	3944	10 AF151976	AF151976 Mus muscu
19	1953.5	39.6	3946	10 AF068628	AF068628 Mus muscu
20	1952.5	39.6	4089	10 AF151971	AF151971 Mus muscu
21	1951	39.5	4034	10 AF151975	AF151975 Mus muscu
22	1948.5	39.5	4006	10 AY078427	AY078427 Mus muscu
23	1947.5	39.5	4149	10 AF151972	AF151972 Mus muscu
24	1917.5	38.9	3897	9 AF156487	AF156487 Homo sapi
25	1806.5	36.6	197652	9 AC012074	AC012074 Homo sapi
26	1784.5	36.2	204534	2 AC118195	AC118195 Mus muscu
27	1784.5	36.2	208874	2 AC111092	AC111092 Mus muscu
28	1769	35.8	2848	5 AF135438	AF135438 Danio rer
29	1624.5	32.9	3017	9 AK001191	AK001191 Homo sapi
30	1504	30.5	1758	9 BC018214	BC018214 Homo sapi
31	1409.5	28.6	210269	2 AC116459	AC116459 Mus muscu
32	1142.5	23.2	185413	2 AC119009	AC119009 Rattus no
33	1006.5	20.4	119630	2 AC112040	AC112040 Rattus no
34	996.5	20.2	167568	2 AC111734	AC111734 Rattus no
35	994	20.1	2127	9 AF129267	AF129267 Homo sapi
36	872.5	17.7	2057	9 AF129268	AF129268 Homo sapi
37	785	15.9	2008	9 AF129269	AF129269 Homo sapi
38	748	15.2	2077	6 AR129189	AR129189 Sequence
39	728.5	14.8	119630	2 AC112040	AC112040 Rattus no
40	693.5	14.1	1397	9 BC002560	BC002560 Homo sapi
41	689	14.0	1705	9 AF194032	AF194032 Homo sapi
42	643.5	13.0	1589	10 AF220524	AF220524 Mus muscu
43	643.5	13.0	1645	10 MM0404467	MM0404467 Mus muscu
44	562.5	11.4	244329	2 AC107644	AC107644 Mus muscu
45	465	9.4	117264	2 AC103454	AC103454 Rattus no

RESULT 1

#### ALIGNMENTS

AF067972 3005 bp mRNA linear PRI 12-FEB-2001  
 LOCUS Homo sapiens DNA cytosine methyltransferase 3 alpha (DNMT3A) mRNA,  
 complete cds.  
 AF067972 10433969  
 AF067972.2 GI:12746531  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 Xie S., Wang Z., Okano M., Nogami M., Li Y., He W.W., Okumura K.  
 1 (bases 1 to 3005)  
 Xie S., Wang Z., Okano M., Nogami M., Li Y., He W.W., Okumura K.  
 and Li E.  
 Cloning, expression and chromosome locations of the human DNMT3  
 gene family  
 Gene 236 (1), 87-95 (1999)  
 JOURNAL MEDLINE  
 PUBMED 99365304  
 10433969  
 REFERENCE 2 (bases 1 to 3005)  
 Xie S., Okano M. and Li E.  
 Direct Submission  
 TITLE Submitted (25-MAY-1998) CVRC, Mass. General Hospital, 149 13th  
 Street, Charlestown, MA 02129, USA  
 JOURNAL 3 (bases 1 to 3005)  
 Hata K., Shirochu H., Saeaki H. and En L.  
 Direct Submission  
 TITLE Submitted (12-FEB-2001) CVRC, Mass. General Hospital, 149 13th  
 Street, Charlestown, MA 02129, USA  
 JOURNAL Sequence update by submitter  
 REMARK On Feb 12, 2001 this sequence version replaced gi:4927369.  
 FEATURES  
 source  
 1..3005  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2p23; between D2S171 and D2S174; 48-50 cM"  
 1..3005  
 /gene="DNMT3A"  
 230..2968  
 /gene="DNMT3A"  
 /functions="de novo DNA cytosine methyltransferase"  
 /codon\_start=1  
 /product="DNA cytosine methyltransferase 3 alpha"  
 /protein\_id="AA03084.2"  
 /db\_xref="GI:12746532"  
 /translation="MPAMPSSPGDTSAAEREDRKGDEEPRGKEERQEPSTT  
 ARKVRGRKRKHPVPSGDTKPAVISKSPSAQDSGASELLPNGLDKERSEPOE  
 EGSFAGKQKGAPEAGEGAEATLPEASRAVNGCCTPKEGRGAPAGKEKRETHES  
 WMEGSGRLGGLGHESSLRQRPMLRTFQNGDPYVLSKRDDEWLASWKEAEKKA  
 KYIAGMANVENQSPGESOKVEASPPAVQPTDPASPTVATTPBPVSGADGKNATK  
 AGDPEFEDGRFGIGELVWKLGFNSWFRIRVSWMTGRSRAEGRWMMFGDG  
 RFSVVCVKLPLSSPCSAFHQATNKQPMIRKAIYEVLOVASRAKGLFPVCHDSDE  
 SDTAKAVEQNKMIEALWGFPQSPGKLEPPBEENKPYEVTDMMWPEAAAYAP  
 PPPAKPRKSTAEKPKVKEIIDERTRELVVEVRKKNIEDICISGSLNVTLEHPL  
 FVFGQCNKNCFLCAYQYDDDDGYSTICCGREVLMCGNCCRCFCVECDLL  
 VPGAAQAATKEDPWCYMGHGKTYGLRRRDPMSRLQMFANNHDDQEPFPKVP  
 YVPAEKRPRLVSLDFGATGLLVKGLGTQVDRTYASEVCEDSITYGVMSHQSIM  
 YVGVRSYTKHQEWGPFDLVIGSPNCNLSIVNPARKGLYEGRGLRFFEFYRLHLD  
 ARPKEGDDRPFFWLFFENVMGVSDKDRSRFLESNPMIDAKEVDSAAHRAFPVFNGL  
 PMNRPLASTVNDLQELCELBHRIAKFSKVRTITRSNSIKQGDQHFPPVFNNEKE  
 DILWCTEMRVFGFPVHYTDVSNMRLAQRLLRLLGSWSVPVIRHPLKVEYFACV"  
 BASE COUNT 674 a 846 c 978 g 507 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1,67e-188 Length: 3005  
 Score: 4935.00 Matches: 912  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-720-086-7 (1-912) x AF067972 (1-3005)  
 QY 1 MetProAlaMetProSerSerGlyProGlyAapThrSerSerSerAlaAlaGluArgGlu 20  
 Db 230 ATGCGCGCCATGCTTCCAGCGCGCCCGGGGACACAGCAGCTCTGCTGGAGCGGAG 289  
 QY 21 GluAapArgLysAapGlyGluGluGluProArgGlyGlyGluGluArgGluGlu 40  
 Db 290 GAGGACCGAAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 349  
 QY 41 ProSerThrAlaArgLysValGlyArgProGlyArgLysArgLysHisProProVal 60  
 Db 350 CCCAGCACCGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 409  
 QY 61 GluSerGlyAapThrProLysAapProAlaValIleSerLysSerProSerMetalGln 80  
 Db 410 GAAAGCGGTGACACGCGCAAGGACCTTGGCGGTGATCTCCAAGTCCCATCGGCCAG 469  
 QY 81 AspSerGlyAlaSerGluLeuLeuProAenGlyAapLeuGluLysArgSerGluProGln 100  
 Db 470 GACTCAGGCGCCTCAGAGCTATTACCAATGGGACTTTGGAGAAAGCGAGTGAAGCC 529  
 QY 101 ProGluGluGlySerProAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 120  
 Db 530 CCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 589  
 QY 121 AlaAlaGluThrLeuProGluAlaSerArgAlaValGluAenGlyGlyGlyGlyGly 140  
 Db 590 GCAGCTGAGACCTTCTGCTGAGGCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAG 649  
 QY 141 GluGlyArgGlyAlaProAlaGluAlaGlyGlyGlyGlyGlyGlyGlyGlyGly 160  
 Db 650 GAGGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 709  
 QY 161 MetLysMetGluGlySerArgGlyArgLeuArgGlyGlyGlyGlyGlyGlyGly 180  
 Db 710 ATGAAATGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 769  
 QY 181 ArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAapProTyrtYrIleSerLys 200  
 Db 770 CGTAGCGCGCATGCGAGGCTCACCTTCCAGCGCGGAGGAGGAGGAGGAGGAGG 829  
 QY 201 ArgLysArgAapGluTptLeuAlaArgTptLysArgGluAlaGluLysAlaLysVal 220  
 Db 830 CGCAAGCGGACGAGTGGCTGGCAGCGCTGGAAAGGAGGAGGAGGAGGAGGAGG 889  
 QY 221 IleAlaGlyMetAsnAlaValGluGluAenGlyGlyGlyGlyGlyGlyGlyGly 240  
 Db 890 ATTGAGGAAATGATGCTGTGGAGAAACACAGGCGCGCGGAGGAGGAGGAGGAG 949  
 QY 241 GluAlaSerProProAlaValGlnGlnProThrAapProAlaSerProThrValAlaThr 260  
 Db 950 GAGGCGCGCTCTCTGTGTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1009  
 QY 261 ThrProGluProValGlySerAapAlaGlyAapLysAenAlaThrLysAlaGlyAapAap 280  
 Db 1010 AGCGCTGAGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1069  
 QY 281 GluProGluTyrGluAapGlyArgGlyPheGlyGlyGlyGlyGlyGlyGlyGly 300  
 Db 1070 GAGCCAGAGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1129  
 QY 301 ArgGlyPheSerTptTptProGlyArgLysValSerTptTptMetThrGlyArgSerArg 320  
 Db 1130 CGGGCGCTTCTCTGTTGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1189  
 QY 321 AlaAlaGluThrArgTptValMetTptPheGlyAapGlyLysPheSerValValCys 340  
 Db 1190 GCAGCTGAAGGACCGCTGGGTGATGTGTTGGAGAGGAGGAGGAGGAGGAGGAGGAG 1249  
 QY 341 ValGluLysLeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAan 360  
 Db 1250 GTTGAGAGCTGATGCGCGCTGAGCTCGTTTTCAGTGGCTTCCACAGGCGCGGAGGAGGAG 1309



Qy	361	LYSGINPROWETTYRAGLYALALLETYGLUVALLEUGINVALALASERSETARGLA	380	Qy	721	LYSGILEUTYRGUGLYTHRGLYARGLEUPHEPHEGLUPHETRYARGLEUHHISAP	740
Db	1310	AAGGAGCCCATGATGACCGCAAGGCATCTACGAGGTCCGTGACAGTGGCCAGACGCGCGC	1369	Db	2390	AAGGCGCTTACAGAGGCGACCTGGCGGCTCTTCTTTAGATTCAACCCCTTCACAT	2449
Qy	381	GLYVLEUPHEPROVALCYEHISASPSEASPGLUSEASPETHRALALYALVALGLU	400	Qy	741	ALAARPROLYSGUGLYAASPAPARPROPHETHPLEUPHEGLUAANVALVALA	760
Db	1370	GGAAGGCTTCCGAGTGTGCACGACGATAGAGTGCACCTGCGCAAGCGGTGGAG	1429	Db	2450	GCGCGGCCCAAGAGGAGATGATCGCCCTTCTTCTGGCTCTTTGGAATGTGGTGGC	2509
Qy	401	VALGINSANLYSPROMETILEGLUTRALALEUGLYPHEGLINPROSEGLYPROLYS	420	Qy	761	METGLYALSERASPLYASAGASPILESERARGPHELEUGLUSEAANPROVALMETILE	780
Db	1430	GTGAGAAACAAAGCCCATGATTTGATGGCCCTGGGGGCTTCCAGCCTTCTGGCCCTAAG	1489	Db	2510	ATGGCGCTTGTGACAAAGAGGACATCTCCGATTTCTCGATGTCMAACCTGTGATGATT	2569
Qy	421	GLYLEUGLUPROPROGLUGLUGLULYSANBPROTYRILYSGILVALTYRTHASPMETTRP	440	Qy	781	ASPALALYSGILVALSERALALAHISARGLAARGTYRPHETTRPGILYANLEUPROGLY	800
Db	1490	GGCTCGAGCCACAGAAAGAGAAAGAAATCCCTACAAAGAGTACACGACGACATGTGG	1549	Db	2570	GATCCCAAGAGTGTGATGCTGCACACAGGCGCCGCTACTTCTGGGGTAACTTCCCGGT	2629
Qy	41	VALGLUPROGLUALALALATYRALAPROPROPROVALALYBLYSPROARGLYSER	460	Qy	801	METASNAPROLEUALASERTHVALASNAPLYLEUGLULEUGLINGULYCYLEUGLU	820
Db	1550	GTGAAACCTGAGGACGCTCTACGACACCTCCACCGCAAAAGCCCGGAGAGAGC	1609	Db	2630	ATGAACAGGCCGTTGGCATCTCATGATGATGAGCTGGAGCTGCAGAGTGTCTGGAG	2689
Qy	461	THRALGLULYSPROLYSVALLYSGULILELEASPIGUAATRGTRARGGLUARGLEUVAL	480	Qy	821	HISGLYARGILALYALYSPHESERYEVALARGTHRIETHRTHARGSERASNSERILE	840
Db	1610	ACAGCGAAGAAAGCCCAAGTCAAGAGATTTATGATGAGCCCAAGAGAGCGCTGGTG	1669	Db	2690	CATGGCAGGANTAGCCCAAGTTCAGCAAGTGAAGACATTACTAGAGGTCAACTCCATA	2749
Qy	481	TYRGLUVALARGLULYSARGASNILEGLUASPILECYSLIESERTCYSGILYSERLEU	500	Qy	841	LYSGINGLYLSAEPGLINHISPHETPROVALPHENETASNGULYSGILASPILEUTRP	860
Db	1670	TACGAGGTGCGGAGAGAGTGCAGAACTTGAGGACATCTGATCTCTGTTGGAGCTTC	1729	Db	2750	AAGCAGGGCAAGACAGCAATTTCTCTGCTTCATGATGAGAAAGAGCATCTTATGG	2809
Qy	501	ASNVALTHRILEUGLULHISPROLEUPHEVALGLYGLMETCYEGLINANCYEGLYBANCYS	520	Qy	861	CYETHRGILMETGLUARGVALPHGLYPHEPROVALHISTYRTHRAPVALSERASMET	880
Db	1730	AATGTTACCCCTGGAACACCCCTCTTCGTTGAGAGAAATGTCCAAACTGCAAGAACGCG	1789	Db	2810	TGCATGAATGGAAGGATTTGTTTCCAGTCCCATGACTGATGAGTCCAAACATG	2869
Qy	521	PHLEUGLUCYSALATYRGLINTYRASPAPAPGLYTRGINSERTYRYSERTHRILEYCS	540	Qy	881	SEARGLEUALAARGLINAARGLEUGLYARGSETRPSEVALPROVALILEARGHIS	900
Db	1790	TTTCTGAGTGTGGTACCACTAGACGACGACGACGCTACCTACTGACCATCTGCG	1849	Db	2870	AGCGCGTTGGGAGGACAGAGCTGCTGGCGCGTCATGAGAGCGTCCAGTATCTGAC	2929
Qy	541	CYSGILYGLYARGLULALLEUMETCYSGLYASNANBANCYSARGCYSPHECYVAL	560	Qy	901	LEUPHEALPROLEULYSGILUTYRPHALACYVAL	912
Db	1850	TGTGGGGCGGTGAGTGTCTCATGTGCGAAACAACACTGTGAGTGTGCGTGTG	1909	Db	2930	CTCTTCGCTCCGCTGAAGAGTATTTGCGTGTG	2965
Qy	561	GLUCYSVALAPLEULEUVALGLYPROGLYALALAGLINALALILEUGLULASPRO	580	RESULT 2			
Db	1910	GAGGTGTGGACCTCTGTGTGGGCGCGGCTGCGCCAGGACCATTAAGAAAGACCCC	1969	AFJ31856	AFJ31856	4258 bp	mRNA
Qy	581	TRPASNCTYRMETCYSGLYHISLYSGILYTHRYGLYLEUARGXARGGLUNAP	600	LOCUS			
Db	1970	TGGAACCTGCTACATGTGCGGACAAAGGATACCTACGGGCTGTGCGGCGGAGAGAC	2029	DEFINITION			
Qy	601	TRPROSEATRGLEUGLIMETPHEPHEALASNSNHSASPGLINGLUPHEASPPROPRO	620	ACCESSION			
Db	2030	TGGCCCTCCGGCTCCAATGTTCTGCTAATAACGACGACGAGATTGACCTCCA	2089	VERSION			
Qy	621	LYSVALTYRPROPROVALPROALAGLULYSARGLYSPROILEARGVALLEUSERLEUPHE	640	KEYWORDS			
Db	2090	AAGGTTTACCCACCTGTCCAGCTGAGAAAGAGAGCCCATCCGGGTCTCTCTCTT	2149	ORGANISM			
Qy	641	ASPGILYLEALATHRGLYLEULEUVALLEULYSAPLEUGLYLEGLVALASPARGYR	660	Homo sapiens.			
Db	2150	GATGGAACTGCTACAGGCGCTCTGTGTGTAAGGCTTGAGGATTCAAGTGCACGCTAC	2209	Homo sapiens			
Qy	661	ILEALASERGLUALCYSGILUASPSERILETHRVALGLIMETVALARGHISGLINGLYS	680	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Db	2210	ATTGCTCGGAGGTGTGTGAGGACTCCATCAAGTGGGCAATGTGTGCGCAACAGGAGAG	2269	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
Qy	681	ILEMETYRVALGLYASPAVALARGSERVALTHRGILYSHISILEGLINGLUTRPGLYPRO	700	REFERENCE			
Db	2270	ATCATGTGTGCTGGGAGCTCGCAGCTGCACAGAAAGCATATCAGAGATGTGGGCCCA	2329	1 (bases 1 to 4258)			
Qy	701	PHASPLEUVALILEGLYGLYSERPROCYASNAPLEUSERILEVALASNPROVALARG	720	2 (bases 1 to 4258)			
Db	2330	TTGCAATGTGTGATGTGGGAGCTCCCTGCAATGACCTCTCATGTCAACCTGTCTGCG	2389	NI, J., Pradhan S. and Roberts, R. J.			
				Direct Submission			
				Submitted (22-DEC-2000)			
				New England Biolabs, 32 Tozer Road,			
				Beverly, MA 01915, USA			
				FEATURES			
				location/Qualifiers			
				1..4258			
				/organism="Homo sapiens"			
				/db_xref="taxon:9606"			
				/chromosome="2"			
				1..4258			
				/gene="DNMT3A"			
				227..2956			
				/gene="DNMT3A"			
				227..2956			
				/codon_start=1			
				/product="DNA cytosine methyltransferase 3 alpha"			
				/protein_id="AAU57039.1"			

Qy	1	MetProAlaMetProSerSerGlyProGluYAspThrSerSerAlaAlaGluArgGlu	20
Db	218	ArgCCCGCCATGCGCTCCACAGCGCCCCGGGGACACACAGAGCTCTGTCGGAGCGGGAG	2778
Qy	21	GluAspArgLysAspGlyGluGluInGluProArgGlyLysGluArgGluGlu	40
Db	278	GAGCAACGAAGACCGAGAGGACGAGAGAGCCGCGCTGGCAAGGAGAGGCGCCAAAG	3378
Qy	41	ProSerThrThrAlaArgLysValGlyArgProGlyArgLysArgLysHisProVal	60
Db	338	CCACGACACCGCACGGAAGGTGGCGCGCTGGAGGAAGCGCAAGCACCCCCCGGTG	3978
Qy	61	GluSerGlyAspThrProLysAspProAlaValIleSerLysSerProSerMetAlaGln	80
Db	398	GAAGCGGTGCACGCGCAAGAGCCCTCGCGTGATCTCCAAGTCCCATCCATGCCCCAG	4578
Qy	81	AspSerGlyAlaSerGluLeuLeuProAsnGlyAspLeuGluLysArgSerGluProGln	100
Db	458	GACTCAGCGCCTCAGAGCTATTACCCAAATGGGGACTTGGAGAAAGCGGAGTGAGCCCCAG	5178
Qy	101	ProGluGluGlySerProAlaGlyGlyGlnLysGlyGlyAlaProAlaGluGlyGluGly	120
Db	518	CCAGAGGAGGGGAGCCCTGCTGGGGGGCAGAGAGGGCGGGGCCCCAGCAGAGGGGAGAGGT	5778
Qy	121	AlaAlaGluThrLeuProGluAlaSerArgAlaValGluAsnGlyCysCysThrProLys	140
Db	578	GCAGCTGAGACCCCTGCCTGAGCCCTCAAGAGCAGCTGGAAAAATGGCTGCTGCACCCCCAAG	6378
Qy	141	GluGlyArgGlyAlaProAlaGluAlaGlyLysGluGlnLysGluThrAsnIleGluSer	160
Db	638	GAGGCCGAGAGACCCCTCGAGNAGCGGGCAAGAAACAGAGAGAGACCAACATCGATCC	6978
Qy	161	MetLysMetGluGlySerArgGlyArgLeuArgGlyGlyLeuGlyTrpGluSerSerLeu	180
Db	698	ATGAAAAATGGAGGGGCTCCCGGGGGCGGCTCGGGGGTGGCTGGGGCTGGAGGTCACGCTC	7578
Qy	181	ArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyrTyrIleSerLys	200
Db	758	CQTCAAGCGGGCCATATCGCAGGCTCACCTTCACGGCGGGGGGACCCCTACTACATCAGCAAG	8178
Qy	201	ArgLysArgAspGluTrpLeuAlaArgTrpLysArgGluAlaGluLysLysAlaLysVal	220
Db	818	CGCAAGCGGACCGAGTGGCTGGCACTCTGAAAGGGAGGCTGAGAGAGAAAGCCAAAGTC	8778

```

Qy 581 TTPASNCSTYMETCYAGLYHISLYSGLYTHRTRYGLYLEUENARAGRLUASP 600
Db 1958 TGGAACTGCTCATGTGTGGGACACAAGGATCTTACGGGGCTGTGGGGGAGAGAC 2017
Qy 601 TTPROSERATGLENGLMETPHEPHEALASNAENHISAPNGINGLUPhASPProPro 620
Db 2018 TGGCCCTCCGGCTCCAAATGTTCTTGGCTAATAACACAGACAGAAATTTGACCTTCCA 2077
Qy 621 LYSVALTRYPROPROVALPROAGLULYSARGLYSPROILEARVALLEUSERLEUPHE 640
Db 2078 AAGGTTTACCACTCGTCCAGAGCTGAGAAAGGAAAGCCATCCGGGTCTGTCTCTTT 2137
Qy 641 ASGLYLLIETATHTGYLEULEUVALLEULYSASPLEUGLYLLEGLINVALASPARTRY 660
Db 2138 GATGAAATCGCTACAGGGCTCTGTGTCTGAAAGGACTTGAGGACTTCAAGTGGACCGCTAC 2197
Qy 661 LLELASERGLUVALCYAGLULASPSERLETHRVALLIGLYMETVALARGHISGLNGLYLS 680
Db 2198 ATGCTCTGGAGGTGTGTGAGACTCTCCATCACAGTGGGCTGTGGGACACAGGGAGAG 2257
Qy 681 LLEMETRYVALGLYASPVALLARGSERVALTHRGINLYSHISLLEGLNGLUTRPGLYPRO 700
Db 2258 ATCATGTACGTCTGGGAGAGTCCGACGCTCACACAGAAACATATCCAGAGTGGGCCCA 2317
Qy 701 PHEASPLEUVALLEGLYGLYSETPROCYASNAASPLEUSERILEVALASPROALLARG 720
Db 2318 TTGCATCTGTGATTTGGGGGACAGTCCCTGCATGACCTTCCATCGTCAACCTCTCGC 2377
Qy 721 LYSGLYLEUTRYGLINGLYTHRGLYARGLEUPHEPHEGLUPHETRYARGLEULENHISAP 740
Db 2378 AAGGGCTCTACGAGGGACAGTGGCGGCTCTTGTGAGTTCTTACCGCTCTCTGCATGAT 2437
Qy 741 ALAARGPROLYSGLYGLYASPARARPPROPHETRPLEUPHEGLUASNVALVALALA 760
Db 2438 GCGGCGCCCAAGAGGAGAGATGATGCCCCCTTCTTGTGAGAAATGGTGGGCC 2497
Qy 761 METGLYVALSERASPLEYARGASPLESERARGPHELEUGLUSERASPROVALMETILE 780
Db 2498 ATGGCGCTTATGACACAGAGGACATCTCGCATTTCTGAGTCCACCCCTGTGATGATT 2557
Qy 781 ASPALALYRGILVALSERIALALAHISARGALAAATGYTRPHETRGLYASPLEUPROGLY 800
Db 2558 GATGCCAAGAAGTGTGAGCTGCACACAGGGCCCGCTACTTGTGGGTAAACCTTCCGGT 2617
Qy 801 METASARGPROLEUALSERETHRVALLASAPLYSLEUGLULEUGLUCYLEUGLU 820
Db 2618 ATGAACAGGCGCTGTGCATCTCATGTAATGATTAAGCTGAGCTGCAGAGTGTCTGGAG 2677
Qy 821 HISGLYARGILEALALYSPHESELYSERVALARGTHRIETHTHARGSERASNSERILE 840
Db 2678 CATGCCAAGATAGCCCAAGTTGACAAAGTGAAGACATTACTACAGAGTCAAACTCCATA 2737
Qy 841 LYSGLINGLYLYASPGLNHISPHETPROVALPHEMETASNGLULYGLUASPLEULEUTRP 860
Db 2738 AAGCGGGCAAGACAGACATTTCTGTCTTCAAGAAATGAGAAAGAGACATCTTATG 2797
Qy 861 CYSATHRGUMETGLUARGVALPHEGLYPHEPROVALHISTYTRHASPVALSERASNET 880
Db 2798 TGCATGAAATGAGAAAGGATATTGTTGTTCCAGTCCATATACGACCTCTCCAAACATG 2857
Qy 881 SERARGLEUALAARGLARGLEULEUGLYARGSETRIPSERVALPROVALLEARGHIS 900
Db 2858 AGCGCTTGGCGAGGACAGAGACTGCTGGGCGGTGATGAGAGTGCACACTATCCGCCAC 2917
Qy 901 LEUPHEALAPROLEULYSGLYTRYPHEALACYVAL 912
Db 2918 CTCTTCGCTCCGCTGAGAGATATTTTGGTGTGTG 2953

```

```

RESULT 3
LOCUS AF068625 4192 bp mRNA linear ROD 06-DEC-1999
DEFINITION Mus musculus DNA cytosine-5 methyltransferase 3A (Dnmt3a) mRNA,
complete cds.

```

```

ACCESSION AF068625
VERSION AF068625.2 GI:6449467
KEYWORDS
SOURCE
ORGANISM Mus musculus.
REFERENCE
AUTHORS Okano, M., Xie, S., and Li, E.
TITLE Cloning and characterization of a family of novel mammalian DNA
(cytosine-5) methyltransferases
JOURNAL Nat. Genet. 19 (3), 219-220 (1998)
MEDLINE 98324766
PUBMED 9662389
REFERENCE
AUTHORS Xie, S., Okano, M., and Li, E.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1998) CYRC, Mass. Gen. Hospital, 149 13th Street,
Charlestown, MA 02129, USA
REFERENCE
AUTHORS Okano, M., Chijiwa, T., Sasaki, H., and Li, E.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-1999) CYRC, Mass. Gen. Hospital, 149 13th Street,
Charlestown, MA 02129, USA
REMARK
COMMENT Sequence update by submitter
FEATURES
On Nov 18, 1999 this sequence version replaced gi:3327977.
location/Qualifiers
1..4192
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="12"
/map="4.0 CM"
1..4192
/gene="Dnmt3a"
1..4192
/gene="Dnmt3a"
/note="similar to EST sequences deposited in GenBank
Accession Numbers AA052791, AA111043, AA154890, AA240794,
AA756653, W58898, W59299, W91664, and W91665"
217..2943
/gene="Dnmt3a"
/function="de novo DNA methylation"
/note="contains Cys-rich region"
/codon_start=1
/product="DNA cytosine-5 methyltransferase 3A"
/protein_id="AAC40177.2"
/db_xref="GI:6449468"
/translation="MPSSGPGDTSSSSLEREDDRKEGEEOENKGEERDEPSATARK
VGRPRGRKHPVESSDTPKDPATVTSQPMADQSGSDLLPNGLERSRPQEGS
PAAGQKGAAGAEGETETPEASRAVENGCCTYKRGASGSEKQOTVIESKME
GSRGRLAGGAGESSLGRORPMPLRTPQAGDPVYISKRDEPLARWRKEAKKAVIA
VMAVENQASGSGQKYEASPRAYOQTPAPSTVAITPEPVGADGADKATYAAD
EPYEDRGRGFIELVNGKLRGFSWVGRIYSWMTGRSRAABGTIRVVMFGDKFSY
VCVEKMLPLSSFCAPHOATYNNQPMYRKAILEVLAQASSPAGLFPACHDSDSG
KAVEVQNKQMIEMALGFOPSGPGLPPEEKPYEVTYDMVAVEBAAYAPPPA
KKPRSTTEKPKYKEIIDERTRERLYVEVOKCNIEDI.CISCSGLVNTLEHPITGG
MCNCKKCFLECAIYOYDGDYOSYCTCCGGEREVLGCGNNCCGCFVCECDLLVPG
AAQALIKEDPMNCTMCGHKSTYGLIRREMPRLQMPFRANHOEDPDRVYVPPA
EKRPKIRIVSLPDLATGLVLKDLGIQVRYLISEVCEDSITVGMVHQQILNYGD
VRSVTOHQIQEWGPFDLVIGSPENDLSIVNPAKGLYEGTGLFFEFRLHDPK
EGDGRPPFVLFENVVAVGVDKDRDISRFLSNPVIDAKESAAHARARYFMGNTPGMN
RPLASTVNDKLEOECHGRIRAKFSVRTITTSNSITKQCKDHPVFMNEKEDILW
CTEMERVGFPPHYHTDYSNNSRLARORLLRSWSVPIRHLFALRKEYFACV"

```

BASE COUNT 1094 a 1072 c 1186 g 840 t

## ALIGNMENT SCORES:

```

Pred. No.: 1,26e-180 Length: 4192
Score: 4740.50 Matches: 877
Percent Similarity: 97.15% Conservative: 8
Best Local Similarity: 96.27% Mismatch: 25
Query Match: 96.06% Indels: 1
DB: 10 Gaps: 1

```



QY 722 G1yleuTYrglunGIyThG1yAArgleuBhepHeG1upheTYrAArgleuLeuH1sAspAla 741  
 Db 2368 GGACTTATGAGGAGTACGGCCGCTCTTCTTGATGCTTACCCGCTCCTGCATGATGCG 2427  
 QY 742 ArgProlyserG1unGIyAAspAArgProBhepHeT1rP1eupHeG1uAspAlaValAlaMet 761  
 Db 2428 CGGCCCAAGGAGGAGATGATCGCCCTTCTTCTGCTTCTTGAAGATGATGCTGCGCATG 2487  
 QY 762 G1yAlaSerAAspAArgAAspAl1eSerAArgPheleuG1uSerAAspProValMet1leAsp 781  
 Db 2488 GGCGTTATGACAGAGAGGACATCTCGCATTTCTTGAGCTTACCCCGTATGATTAAC 2547  
 QY 782 A1aArgG1uValSerAlaAlaH1sAArg1aAArgTYrPheT1rP1eupHeG1uAspProG1yMet 801  
 Db 2548 GCCAAGAAGTGTCTGCTGCACACAGGCGCTTACTTCTGGGGTAACCTTCTGCGCATG 2607  
 QY 902 AsnAArgProLeuAlaSerThValAsnAAspAArgleuG1uLeuG1uCySeuG1uH1s 821  
 Db 2608 AACAGGCTTGGCATCCACTGTGATGATTAAGCTGAGCTGCAAGAGTGTCTGAGACAC 2667  
 QY 822 G1yAArg1eAlaAArgPheSerAArgThValAArgTh1eThThAArgSerAAspSer1leAsp 841  
 Db 2668 GGCAAGATAGCCAGATTCAGACAAAGTGAAGACATTACCAAGCTCAAACTTAATAAG 2727  
 QY 842 G1unGIyAAspAArg1uH1sPheAArgProValPheMetAAspG1uAAsp1leuAArgProCys 861  
 Db 2728 CAGGCAAGAAGACACGATTTCCCTCCCTCTTCATGAACGAGAGAGACATCTGTGTGTC 2787  
 QY 862 ThG1uMetG1uAArgValPheG1yPheProValH1sTYrThAAspValSerAAspMetSer 881  
 Db 2788 ACTGAAAGGAGGAGTGTGTGCTTCCCTCCACTACACAGAGCTGTCTCCAACTGAGC 2847  
 QY 882 Arg1eAlaAArgG1uAArgleuG1uAArgSerAArgSerAArgProVal1leAArg1eLeu 901  
 Db 2848 CGCTTGGGAGGACAGAGCTGTGAGCCGATGTGAGAGGTGCGGATGTCCGCAACCTC 2907  
 QY 902 PheAlaProLeuAArgG1uTYrPheAlaCysVal 912  
 Db 2908 TTCGCTCCGCTGAAGATATTTTGTGTGTG 2940

RESULT 4  
 BC007466 4094 bp mRNA linear ROD 07-AUG-2002  
 LOCUS Mus musculus, DNA methyltransferase 3A, clone MGC:5662  
 DEFINITION IMAGE:3492853, mRNA, complete cds.  
 ACION BC007466  
 VERSION BC007466.1 GI:13938620  
 KEYWORDS MGC.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.

REFERENCE  
 AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 TITLE Strauberg, R.  
 JOURNAL Direct Submission  
 Submitted (01-MAY-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [gsapbs-remail.nih.gov](mailto:gsapbs-remail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (ILMT)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mdedpaxil.stanford.edu](mailto:mdedpaxil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
 R. M.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/ILMT at: <http://image.llnl.gov>  
 Series: IRAC Plate: 8 Row: n Column: 11  
 This clone was selected for full-length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 6681208.

## FEATURES

## Source

1. 4094  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /map="C2ECH 11"  
 /clone="MGC:5662 IMAGE:3492853"  
 /tissue\_type="Mammary tumor metastatized to lung. Tumor  
 arose spontaneously from a senescent normal mammary  
 (clonal) outgrowth infected with the virus MMTV."  
 /clone\_1ib="NCI CGAP\_Lu29"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 245. 2971  
 /codon\_start=1  
 /product="DNA methyltransferase 3A"  
 /protein\_id="AAH07466.1"  
 /db\_xref="GI:13938621"  
 /db\_xref="locusid:13435"  
 /translation="MSSSGFDITSSSLEREDRKSGEEDENRKRGEPPATARK  
 VGRPRGRKHPVSSDTPDPVATTKSQPMADSGSDLLPMDLEKRESEPPDESS  
 PAAGQKGAFAEGEGTETPEASRAVENGCCTVKEGAGAGAEKEKPNIESMKV  
 GSGRLRGAGMSLRLRPMPLTPGAGPYYSIKRDEMLRMREBAKKAVIA  
 VMVAENRNASGSGOKVEASPAVQDTPAPATVATPEPQVGNATQAAAD  
 EPYEDRGRGIGELVNGKRGFSWMPGRVSMNTRGSAABETRRVMGRDQKRSV  
 VCEKMLPLSFCSAFHQATYNNKQPMRKATLYELVQVASSRAGLFPACDSDSDSG  
 KAVEQNKOMIEMALGGFQSGPGLPEEKNRPKEVYTTDMVVEEALAAVPPA  
 KKRRKSTTERPKYKEIIDERTRELVVEVEOKCRNIEDICISCGSLVTTVEPLFIG  
 MCONKRCFLCEAYDDGYOYCTCCGREGVLMGNNCCSPCEVCVDLVPFG  
 AAOAIKEDPMNCGMCGKGTGLRRRDMPSRLQWEPANNHDOERDPPVYPVPA  
 EKKPRIVSLPFGIATGLVLKDLGIQVRYIASBCEDSITGMWRHCKIMYVD  
 VRSTQTHIEMGFPDLVIGSSPCNDISIVNPAKGLVEGTGRLEFETRLHDARK  
 EGDRPPEFLFENVVAMGVSdkRD1SFLPSNPVMDIAKVSAAHRAARYFMGNLPGNN  
 RPLASTVNDKLELOECLEHGR1AKFSKVRTITTRNSN1IKQKDHFPVFMNEXKDIILW  
 CTEWERVYVGPVYHTDVSNMSRLAROLLRSMNSVPVIRHLFAPLKEYFACV"

## CDS

BASE COUNT 1067 a 1057 c 1164 g 806 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.63e-180 Length: 4094  
 Score: 4737.50 Matches: 877  
 Percent Similarity: 97.04% Conserves: 26  
 Best Local Similarity: 96.27% Mismatches: 26  
 Query Match: 96.00% Indels: 1  
 DB: Gaps: 10

US-09-720-086-7 (1-912) x BC007466 (1-4094)

QY 2 ProAlaMetProSerSerG1yProG1yAAspThrSerSerA1a1eG1uAArgG1uG1u 21  
 Db 239 CCAGCAATGACCTCCAGCGGCGCCGAGACACAGACACTCTCTCTGAGGAGGAGAT 298  
 QY 22 AspAArgAAspG1yG1uG1uG1uG1uProAArgG1yAAspG1uAArgG1uG1uPro 41  
 Db 299 GATCAAG 358  
 QY 42 SerThThAlaAArgAArgValA1yAArgProG1yAArgAArgAArgAArgProValG1u 61  
 Db 359 AGCGCAAGGCGCCGGAAG 418  
 QY 62 SerG1yAAspThrProAAspProAlaVal1leSerAArgProSerMetAlaG1uAsp 81  
 Db 419 AGCACTGACACCCCAAG 478  
 QY 82 SerG1yAAspThrProAAspProAAspProAAspProAAspProAAspProAAspPro 101  
 Db 479 TCTGACCCCTCAGATCTCTACCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 538  
 QY 102 G1uG1uG1ySerProAlaG1yG1uG1yG1yA1aProAlaG1uG1yG1uG1yA1a 121

Db 539 GAGGAGGAGCCAGCTGCAAGGCGAGAAAGGCTGGGGCCCAAGCTCAAGAGAGAGGA --- 595  
Qy 122 AlaGluThrLeuProGluAlaSerArgAlaValGluAsnGlyCysCysThrProLysGlu 141  
Db 596 ACTGAGACCCCAAGAGCTCTCAGAGCTGTGGAGATGGCTGTGTGACCAAGGAA 655  
Qy 142 GlyArgGlyValProAlaGluAlaGlyValGluGluLysGluThrAsnGluSerMet 161  
Db 656 GGCCTGTGAGCTCTGCGAGAGAGGCGAAGAACAGACAGACCCCAACATCGATCCATG 715  
Qy 162 LysMetGluGlySerArgGlyArgLeuArgGlyGlyLeuGlyTTPGluSerSerLeuArg 181  
Db 716 AAATGGAGGCTCTCCGGGGCCGACTGCGAGGTGGCTTGGCTGGGAGTCCAGGCTCCGT 775  
Qy 182 GlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyrTyrIleSerLysArg 201  
Db 776 CAGCGACCATCCCAAGACTCACCTTCAGCGACGGGACCCCTACTACATCAGCAACGG 835  
Qy 202 LysArgAspGluTTPLeuAlaArgTTPLysArgGluAlaGluLysLysAlaLysValIle 221  
Db 836 AAACGGGATGAGTGGCTGGCACGTTGGAAAGGGAGGCTGAGAAAGAAAGCCCAAGGTAAAT 895  
Qy 222 AlaGlyMetAsnAlaValGluGluAsnGlnGlyProGlyGluSerGlnLysValGluGlu 241  
Db 896 GCAGTAATGAATGCTGTGGAAGAGAACCCAGGCTCTGGAGAGTCTCAGAGGTGGAGAG 955  
Qy 242 AlaSerProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrThr 261  
Db 956 GCCAGGCTCTCTGCTGTCAGACAGCCCAAGGACCTTCTCCGACTGTGGCCACACCC 1015  
Qy 262 ProGluProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAspGlu 281  
Db 1016 CTTGAGCCAGTAGAGGGATGCTGGGGACAAAGATGCTACCAAGACGCCAGCATGAG 1075  
Qy 282 ProGluTyrGluAspGlyArgGlyPheGlyIleGlyGluLeuValTTPGlyLysLeuArg 301  
Db 1076 CTTGAGTAGAGATGGCCGGGCTTTGGCAATTTGGAGACTGCTGTGGGGAAATCTTCGG 1135  
Qy 302 GlyPheSerTTPTTPProGlyArgIleValSerTTPTTPMetThrGlyArgSerArgAla 321  
Db 1136 GGCTTCTCTGTCGGCCAGGCGGAATTTGTGCTTGTGGTAGATGGCAAGTTCTCAGTGGTGTGTG 1195  
Qy 322 AlaGluGlyThrArgTTPValMetTTPPheGlyAspGlyLysPheSerValValCysVal 341  
Db 1196 GCTGAAGGCACTCGCTGGGTCTATGTTGTCGAGATGGCAAGTTCTCAGTGGTGTGTGTG 1255  
Qy 342 GluLysLeuMetProLysSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLys 361  
Db 1256 GAGAAGCTCATCCGCTCAGCTCCTTCTGCAGTGCATTCACAGGCCACCTACACAG 1315  
Qy 362 GlnProMetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGly 381  
Db 1316 CAGCCATGTACCCGAAAGCCATCTACAAAGTCTCTCAGTGGCCAGCAGCCGTCGCCGG 1375  
Qy 382 LysLeuPheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValGluVal 401  
Db 1376 AAGCTGTTTCCAGCTTGCCATGACAGATGAAAGTGAAGTGAAGGCAAGGCTGTGGAAAGTG 1435  
Qy 402 GlnAsnLysProMetIleGluTTPAlaLeuGlyGlyPheGlnProSerGlyProLysGly 421  
Db 1436 CAGAAACAGACAGATGATTAATGGGCCCTCGGTGGCTTCCAGCCCTCGGGTCTTAAGGGC 1495  
Qy 422 LeuGluProProGluGluLysAsnProTyrLysGluValTyrThrAspMetTTPVal 441  
Db 1496 CTGGAACCAACAGAGAGAGAGAAAGATCTTACAGGAAGATTTACCCGACATGTGGGTG 1555  
Qy 442 GluProGluAlaAlaIleTyrAlaProProProAlaLysLysProArgLysSerThr 461  
Db 1556 GAGCCTGAAGCAGCTGCTTACGCCCCCAACCCAGCCAGCAAGAAACCCAGAAAGAGCACA 1615  
Qy 462 AlaGluLysProLysValLysGluIleLeuAspGluArgThrArgGluArgLeuValTyr 481

Db 1616 ACAGAGAAACCTAAGGTCAAGGAGATCATTTGATGAGCGCACAGGGAGCGCTGGTGAT 1675  
Qy 482 GluValArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsn 501  
Db 1676 GAGGTGGCCAGAGAGTGCAGAAACATTCGAGGACATTTGTATCTCATGTGGGAGCTCAAT 1735  
Qy 502 ValThrLeuGluHisProLeuPheValGlyGlyMetCysGlnAsnCysLysAsnCysPhe 521  
Db 1736 GTACCCCTGGAGACCCACTCTTCATTTGGAGCATGTGCCAGAACTGTAAGAACTGCTTC 1795  
Qy 522 LeuGluCysAlaTyrGlnTyrAspAspGlyTyrGlnSerTyrCysThrIleCysCys 541  
Db 1796 TTGGAGTGTGCTTACCAAGTATGACGAGATGGGTACCAAGTCTTATTGCACCATCTGCTGT 1855  
Qy 542 GlyGlyArgGluValLeuMetCysGlyAsnAsnCysCysArgCysPheCysValGlu 561  
Db 1856 GGGGGGCGCTGAAGTGTCTATGTGTGGGAACAACACTGCTCAGGCAGCATTAAGGAAGACCCCTGG 1915  
Qy 562 CysValAspLeuValGlyProGlyAlaAlaGlnAlaIleLysGluAspProTTP 581  
Db 1916 TGTGTGATCTCTTGGTGGGCGCAGGAGCTGCTCAGGCAGCATTAAGGAAGACCCCTGG 1975  
Qy 582 AsnCysTyrMetCysGlyHisLysGlyThrTyrGlyLeuLeuArgArgATGluAspTTP 601  
Db 1976 AACTGCTACATGTGGGGCATTAAGGCACCTATGGCTGCTGCAGACGGGAAGACTGG 2035  
Qy 602 ProSerArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProLys 621  
Db 2036 CCTTCTCGACTCCAGATGTTCTTGGCAATACCATGACCAAGAAATTGACCCCCAAG 2095  
Qy 622 ValTyrProProValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAsp 641  
Db 2096 GTTATACCCACCTGTCGAGCTGAGAGAGAGAGGCCATTCGCGTGTCTCTCTTTGAT 2155  
Qy 642 GlyIleAlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIle 661  
Db 2156 GGGATGCTACAGGCTCTCGTGTGTAAGGACCTGGGCAATCCAAGTGGACCGCTACATT 2215  
Qy 662 AlaSerGluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIle 681  
Db 2216 GCCTCCGAGGTGTGAGGACTCCATCACGCTGGGCATGCTGGCGCACCAAGGAAAGATC 2275  
Qy 682 MetTyrValGlyAspValArgSerValThrGlnLysHisIleGlnGluTTPGlyProPhe 701  
Db 2276 ATGTACGTCGGGACCTCCGAGCTCACAGAAAGCATATCCAGGAGTGGGGCCCATTC 2335  
Qy 702 AspLeuValIleGlyGlySerProCysAsnAspLeuSerIleValAsnProAlaArgLys 721  
Db 2336 GACCTGGTGTATGGAGGAGTCCCTGCAATGACCTCTCCATTGTCAACCTGCGCCCAAG 2395  
Qy 722 GlyLeuTyrGluGlyThrGlyArgLeuPheGluPheThrArgLeuLeuHisAspAla 741  
Db 2396 GGACTTTATGAGGTACTGGCCGCTCTTCTTTGAGTCTACCGCTCTCTGATGATGCG 2455  
Qy 742 ArgProLysGluGlyAspAspArgProPheTTPLeuPheGluAsnValValAlaMet 761  
Db 2456 CGGCCCCAAGGAGGAGATGATCGCCCTTCTTCTGGCTCTTTGAGATGTGGTGGCCATG 2515  
Qy 762 GlyValSerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAsp 781  
Db 2516 GCGCTGTAGTGACAAAGAGGACATCTCGCGATTTCTTGGTCTAAACCCGCTGATGATGAC 2575  
Qy 782 AlalysGluValSerAlaAlaHisArgAlaArgTyrPheTTPGlyAsnLeuProGlyMet 801  
Db 2576 GCCAAAGAGTGTCTGTCACACAGGCGCCGTTACTTCTGGGGTAACCTTCTGGCATG 2635  
Qy 802 AsnArgProLeuAlaSerThrValAsnAspLysLeuGluGlnGluCysValLeuGluHis 821  
Db 2636 AACAGGCTTTGGCATCCACTGTGAATGATTAAGCTGGAGCTGCCAAGAGTGTCTGGAGCNC 2695  
Qy 822 GlyArgIleAlaLysPheSerLysValArgThrIleThrArgSerSerSerIleLys 841  
Db 2696 GCGAGATAGCCAAAGTTTACGAAAGAGGACCATTTACCAAGGCTCAAACTCTATAAG 2755

Qy 842 GlnGlyLysAspGlnHisPheProValPheMetLeuGluValGluAspIleLeuTrpCys 861  
 Db 2756 CAGGCAAGACGACGATTTCCCTCTTCATGAAAGAGAGAGACATCTGCTGTC 2815  
 Qy 862 ThrGluMetGluArgValPheGlyPheProValHisTyrThrAspValSerAspMetSer 881  
 Db 2816 ACTAAATGAGAAAGGGTGTGGCTTCCCTCCCTCACTACAGAGAGCTTCCAACTGAGC 2875  
 Qy 882 ArgLeuAlaArgGlnArgLeuLeuGluValArgSerTrpSerValProValIleArgHisLeu 901  
 Db 2876 CGCTTGGAGAGAGAGAGCTGCTGCGCGATCGTGGAGACCGTCCGCTATCCGACACTC 2935  
 Qy 902 PheAlaProLeuLeuGluValTyrPheAlaCysVal 912  
 Db 2936 TTCGCTCGCTGAGAGATATTGCTTGTGTG 2968  
 RE 5  
 AC120824 123936 bp DNA linear HTG 23-JUL-2002  
 LOCUS Rattus norvegicus clone CH230-282D16, \*\*\* SEQUENCING IN PROGRESS  
 DEFINITION \*\*\* 38 unordered pieces.  
 AC120824 2 GI:21902541  
 VERSION HTG: HTG5\_PHA5E1.  
 KEYWORDS Rattus norvegicus.  
 SOURCE Rattus norvegicus.  
 ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 1 (bases 1 to 123936)  
 Mziny,D.M., Adams,C., Adlo-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amarantunga,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhaq,C., Butch,P., Burkett,C., Burrell,K.D., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P., Gabriel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Hollway,C., Hollins,B., Homsl,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratoch,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Lonsdale,H., Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinec,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokweto,S., Ogun,M., Okwundu,G., Otaguine,N., Ovielo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pyun,L., Qulles,M., Ren,Y., Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sison,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Swalek,A., Tabor,P., Tameris,A., Tameris,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Teitlor,B., Thomas,S., Thomas,S., Umanu,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D., Weinstock,G. and Gibbs,R.  
 TITLE Direct Submission  
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 123936)  
 AUTHORS Worley,K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 123936)  
 AUTHORS Worley,K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Jul 19, 2002 this sequence version replaced gi:20531783.  
 COMMENT  
 -----  
 Center: Baylor College of Medicine  
 Genome Center  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 -----  
 Center project name: GWIV  
 Project Information  
 Center clone name: CH230-282D16  
 -----  
 Summary Statistics  
 Sequencing vector: plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 95907 bases at least Q40  
 Consensus quality: 99223 bases at least Q30  
 Consensus quality: 101036 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 38 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1  
 1009: contig of 1009 bp in length  
 1010: gap of unknown length  
 1110: contig of 1004 bp in length  
 1214: gap of unknown length  
 2214: contig of 1020 bp in length  
 3234: gap of unknown length  
 3334: contig of 1109 bp in length  
 3334: gap of unknown length  
 4443: gap of unknown length  
 4543: contig of 1660 bp in length  
 6203: gap of unknown length  
 6302: contig of 1162 bp in length  
 7464: gap of unknown length  
 7564: contig of 1208 bp in length  
 8772: gap of unknown length  
 8773: gap of unknown length  
 8872: contig of 1499 bp in length  
 8873: gap of unknown length  
 10371: contig of 1243 bp in length  
 10472: gap of unknown length  
 11714: contig of 1110 bp in length  
 11814: gap of unknown length  
 11815: contig of 1110 bp in length  
 11815: gap of unknown length  
 12924: contig of 1323 bp in length  
 12925: gap of unknown length  
 13024: contig of 1205 bp in length  
 14347: gap of unknown length  
 14348: contig of 1205 bp in length  
 15552: gap of unknown length  
 15553: contig of 2044 bp in length  
 17796: gap of unknown length  
 17797: gap of unknown length  
 17896: contig of 1041 bp in length  
 18937: gap of unknown length  
 19037: contig of 2506 bp in length  
 21543: gap of unknown length  
 21544: contig of 2897 bp in length  
 21544: gap of unknown length  
 24540: gap of unknown length  
 24541: gap of unknown length  
 24640: contig of 3472 bp in length  
 28112: gap of unknown length  
 28113: gap of unknown length







Qy 646 G1yleuValleuValysAspieuGlytleGlnValAspaArgTyrIleAlaSerGluVal 665  
Db 86062 GGGCTCCCTGGTGTGAAGACCTGGGACATCCAGTGAGACCCCTACATGCTTCCAGAGGTG 86121  
Qy 666 CysGluAspSerIleThValGlyMetValArgHisGlnGlyValIleMetCysValGly 685  
Db 86122 TGCAGAGCTCATCAACCGTGGCATAGTGGCCACGAGGAAGATATATGTCCTCGGG 86181  
Qy 686 AspValArgSerValThGlnHisIleGlnGluTyrGlyProPheAspLeuValIle 705  
Db 86182 GACGCTGTGCACATCAACACAGAGATATCCAGAGAGGGGCCCTACATCTGGTGT 86241  
Qy 706 G1yGlySerProCysAsnAspieuSerIleValAsnProAlaArgGlyLeuTyrGlu 725  
Db 86242 GGGGCGACGCTCTTGCAGACCTCTCCATCATCAACCTGCCCGGAGGACTTACAG 86301  
Qy 726 G1yThrGlyArgLeuPheGluPheThyArgLeuLeuHisAspAlaArgProGly 745  
Db 86302 GGCACCTGGACCTCTTCTTTGAGTTCTTACCACTCTGCATGATGCGGGCCCAAGAG 86361  
Qy 746 G1yAspAspArgProPhePheThyLeuPheGlu 756  
Db 86362 GGAGAGATCGCCCTTCTTCTGCTCTTGG 86394  
RESULT 6 AK025230 2191 bp mRNA linear PRI 29-SEP-2000  
LOCUS Homo sapiens cDNA: FLJ21577 fis, clone COU06724, highly similar to  
DEFINITION AF067972 Homo sapiens DNA cytosine methyltransferase 3 alpha  
(DNMT3A) mRNA.  
ACCESSION AK025230  
VERSION AK025230.1 GI:10437699  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens colon cDNA to mRNA, clone\_lib:COL clone:COU06724.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (sites)  
REFERENCE Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y.,  
AUTHORS Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T.,  
Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.  
NEBO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 2191)  
Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,  
TITLE Shibahara, T., Tanaka, T. and Nakamura, Y.  
JOURNAL Direct Submission  
REFERENCE Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
JOURNAL University of Tokyo, Laboratory of Genome Structure Analysis, Human  
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,  
Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
FAX:81-3-5449-5416)  
COMMENT NEBO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing: Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing: Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).  
FEATURES  
Source  
1. .2191  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="COL06724"  
/tissue\_type="colon"  
/clone\_lib="COL"  
/note="cloning vector pM18SFLJ3"  
misc\_feature  
1. .2191  
/note="highly similar to AF067972 Homo sapiens DNA  
cytosine methyltransferase 3 alpha (DNMT3A) mRNA"  
BASE COUNT 571 a 505 c 575 g 540 t  
ORIGIN

Alignment Scores:  
Pred. No.: 3.2e-88 Length: 2191  
Score: 2413.00 Matches: 441  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.77% Mismatches: 0  
Query Match: 48.90% Indels: 0  
DB: 9 Gaps: 0  
US-09-720-086-7 (1-912) x AK025230 (1-2191)  
Qy 471 TLeAspGluArgThrArgGluArgLeuValTyrGluValArgGlnIleCysArgAsnIle 490  
Db 2 ATTATAGGCGCAAGAGAGCGGCTGTGTATCAAGGTGGCGGAGAGTGGCGAAGCAT 61  
Qy 491 G1uAspIleCysIleSerCysGlySerIleAsnValThrLeuGlnHisProLeuPheVal 510  
Db 62 GAGACATCTGCATCTCTGTGGAGCCTCAATTTACCTGGAAACCCCTTGTGTT 121  
Qy 511 G1yGlyMetCysGlnAsnCysIleAsnCysPheLeuGluCysAlaTyrGlnTyrAspAsp 530  
Db 122 GGAGGAATGTGCCAAAAGTGCAGAACTGCTTTGTGAGTGTGGCTACAGTACGACGAC 181  
Qy 531 AspGlyTyrGlnSerTyrCysThrIleCysCysGlyGlyArgGluValLeuMetCysGly 550  
Db 182 GACGCTACACAGTCTCTACTGCACCATCTGCTGTGGGGCGGTGAGTGTCTATGTGGGA 241  
Qy 551 AsnAsnAsnCysCysArgCysPheCysValGluCysValAspLeuLeuValGlyProGly 570  
Db 242 AACACAACTGTCTCAGAGTGTGTTGTGGTGAAGTGTGGACCTTTGTGGGGCGGG 301  
Qy 571 A1a1aGlnAla1a1a1leGlyGluAspProTyrAsnCysTyrMetCysGlyHisGly 590  
Db 302 GCTCCCAAGAGCCATTTAAGAGAGACCCCTGGAACCTGACATGATCGGSCACAAGGT 361  
Qy 591 ThrTyrGlyLeuLeuArgArgArgGluAspTyrProSerArgLeuGlnMetPhePheAla 610  
Db 362 ACCTACGGGCTGTGGCGGGCGGAGAGACTGCGCTCCGCGCTCCAGATGTTCTTCCG 421  
Qy 611 AsnAsnHisAspGlnGlnupheAspProProGlyValTyrProProValProAlaGly 630  
Db 422 AATAACCAAGCAGCAAGAAATTTGACCTCCCAAGGTTTACCACTGTGCCAGCTGAGAG 481  
Qy 631 ArgTyrProIleArgValIleuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeu 650  
Db 482 AGGAGGCCATCCGGGTGTCTCTCTTTGATGGAATCGTTACAGGGCTCTGTGTGTG 541  
Qy 651 LysAspLeuGlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIle 670  
Db 542 AAGGACTTGGCATTTCAAGTGCACCGCTACATTCCTCGAGAGTGTGAGGACTCCATC 601  
Qy 671 ThrValGlyMetValArgHisGlnGlyValIleMetTyrValGlyAspValArgSerVal 690  
Db 602 ACGGTGGCGAGTGGCGGACAGGAGAAATCATGACCTCGGGGAGCTCCGAGCGTC 661  
Qy 691 ThrGlnHisIleGlnGlnTyrGlyProPheAspLeuValIleGlyGlySerProCys 710  
Db 662 ACACAGAGCATATCCAGAGTGGGGCCCATTCATCTGGTGAATTTGGGGCAGTCCCTGC 721  
Qy 711 AsnAspLeuSerIleValAsnProAlaArgIleCysGlyLeuTyrGlnGlyThrGlyArgLeu 730  
Db 722 AATGACCTCTCATCGTCAACCTGCTCGCAAGGGCTCTACGAGGGGACTGCGCGGCTC 781  
Qy 731 PhePheGluPheTyrArgLeuLeuHisAspAlaArgProGlyGluGlyAspAspArgPro 750  
Db 782 TTCCTTGTGTTCTACCGCTCTCTGCATATGCGGGCCCAAGAGGAGATGATCGGCC 841  
Qy 751 PhePheTyrLeuPheGluAsnValAlaIleMetGlyValSerAspLysArgAspIleSer 770  
Db 842 TTCCTTGGCTCTTGAAGATGTGTGGCCATGGGCTTACTGTGCAAGAGGAGATCTCG 901  
Qy 771 ArgPheLeuGluSerAsnProValMetIleAspAlaLysGluValSerAlaAlaHisArg 790



```

Db 768 GCTACCAAGTCTCGAGACGTCGA----- 791
Qy 216 LysLysValAlaValIleAlaGlyMetAsnAlaValGluAsnGlnGlyProGlyGlu 235
Db 792 -----GCATCGCTTCAGCAGACAGCCAGTCATCCCTCCAGC 833
Qy 236 SerGlnValGluGlnAlaSerProProAlaValGlnGlnProThrAspProAlaSer 255
Db 834 GTCGACCTTCATGAGAAAGTGCATGAAGCGTCAGTACCCCA-----TCA 881
Qy 256 ProThrValAlaThrThrProGluProValGlySerAspAlaGlyAspLysAsnAlaThr 275
Db 882 GTGACCTGACCCAGATGAGATGAGAGGATGATGATACCAACAGGTGATGACAG 941
Qy 276 LysAlaGlyAspAspGlyProGlyLysLeuAspGlyLysGlyPheGlyIleGlyLeu 295
Db 942 AGCATATATGAGACAGACAGACAGATGATCAGATGATTAAGATTGGAATGATGATCCTC 1001
Qy 296 ValTrpGlyLysLeuArgGlyPheSerTrpTrpProGlyArgIleValSerTrpTrpMet 315
Db 1002 GTGTTGGGAAAGATCAAGGGCTTCTCTGTTGGCTGCGCATGCTGCTGCTGAAAGCC 1061
Qy 316 ThrGlyArgSerArgAlaAlaGlnGlyThrArgTrpValMetTrpPheGlyAspGlyLys 335
Db 1062 ACCTCGAAGCCAGCAGGCGATGCCGAAATGCCCTGGTACAGTGGTTGGTATGGCAAG 1121
Qy 336 PheSerValAlaCysValAlaGlyLysLeuMetProLeuSerSerPheCysSerAlaPheHis 355
Db 1122 TTTTGTGAGATCTCTGCTGACAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1181
Qy 356 GlnAlaThrTyraAsnLysGlnProMetTyraGlyAlaIleTyraGluValLeuGlnVal 375
Db 1182 CTGGTACCTTCATTAATACCTGGTTCTTATAGAGAGCCATGTACCACTCTGAGAAA 1241
Qy 376 AlaSerSerArgAlaGlyLysLeuPheProValCysHisAspSerAspGluSerAspThr 395
Db 1242 GCCAGGTTCCAGCTGCGAAGACCTTC-----TCCGCGAGTCTCT 1280
Qy 396 AlaLysValAlaGluValAlaGlnLysProMetIleGluTrpAlaLeuGlyValPheGln 415
Db 1281 GGAAGTACCTGAGAGACAGCAGTGAAGCCCATGCTGAGTGGGCCAGCGTGGCTTCAAG 1340
Qy 416 ProSerGlyProLysGlyLeuGlnProProGluGlnGlu---LysAsnProTyraLysGlu 434
Db 1341 CCTACTGGAGTGAAGGCTCTCAACCAACAAGAGCAACCCGAAACAAGTGAAGA 1400
Qy 435 ValTyraThrAspMetTrpValGluProGluAlaAlaIleTyraIleProProProAla 454
Db 1401 CGCAACCAAT-----GACTCTGCTGCTTGAAGTCCCGCCCAACCAAG 1445
Qy 455 LysLysProAlaGlySerThrAlaGluLysProLysValLysGluIleLeuAspLysArg 474
Db 1446 CGCTCAAGCAATAATAGCTATGCGCGGAAGAGCGAGGGAG-----GATAGAGAG 1496
Qy 475 ThrArgGluArgLeuValTyraGluValArgGlnLysCysArgAsnIleGluAspIleCys 494
Db 1497 AGCGAGAAACGAGTGGCTTCTCAACTCAACCAACAAGAGGAACTCTGGAAGACCGCTCT 1556
Qy 495 IleSerCysGlySerLeuAsnValThrLeuGluHisProLeuPheValGlyLysCys 514
Db 1557 TTGTCTCTGGAAGAAACAACCTGTCTTCCACCCCTCTTGAAGGTTGGCTCTGT 1616
Qy 515 GlnAsnCysLysAsnCysPheLeuGluCysAlaTyraGlnTyraAspAspAspGlyTyraGln 534
Db 1617 CAGAGTTCGCGGAGTCTCTTCAAGCTCTTCAATGATGATGAGAGCGCTATAG 1676
Qy 535 SerTyraCysThrIleCysCysGlyLysArgGluValIleMetCysGlyAsnAsnAsnCys 554
Db 1677 TCCTACTGACCGTGTGCTGTAGAGCGCTGAACGCTGCTGTGCAAGTGAACCAAGCTGC 1736
Qy 555 CysAlaGlyCysPheCysValGluCysValAspLeuLeuValGlyProGlyAlaAlaGlnAla 574

```

```

Db 1737 TGCAGATCTTCTGTGTGAGATGCTCTGAGAGTCTGCTGGGCGCAGCAGCTGAGCAT 1796
Qy 575 AlaIleLysGluAspProTyraAsnCysTyraMetCysGlyHisLysGlyThrTyraGlyLeu 594
Db 1797 GCCAAGCTGCAAGAAACCTTGAGCGCTATATGTGCTCCCTCAGCGCTGCGATGGGCTC 1856
Qy 595 LeuArgArgArgGluAspTrpProSerArgLeuGlnMetPhePheAlaAsnAsnHisAsp 614
Db 1857 CTCGACGCGCAAGAAAGATTGGACATGCGCTGCAAGACTTCTTCACTACATCTCTAC 1916
Qy 615 ---GlnGluPheAspProProLysValTyraProProValProAlaGluLysArgLysPro 633
Db 1917 CTGAAGAAATTTGAGCCACCCCAAGTTGATCCAGCAATCTTGCAGCCCAAGAGGGGCC 1976
Qy 634 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeuLysAspLeu 653
Db 1977 ATTAGAGCTCTCTCTCTTGTGATGGAATTCACACGGGTACTTGCTGCTCAAGAGTTG 2036
Qy 654 GlyIleGlnValAspArgTyraIleAlaSerGluValCysGluAspSerIleThrValGly 673
Db 2037 GGATTTAAAGTGAAGAAAGTATACATTGCTCCGAAGCTGTGACAGATCCATGCTGTGGGA 2096
Qy 674 MetValArgHisGlnGlyLysIleMetTyraValGlyAspValArgSerValThrGlnLys 693
Db 2097 ACTGTTAAGCATGAAGGCCAGATCAATATATGTCATGACGTCGGAATAATCAAGAAA 2156
Qy 694 HisIleGlnGluTrpGlyProPheAspLeuValIleGlyLysSerProCysAsnAspLeu 713
Db 2157 AATATTGAAGAGTGGGCGCTTGCATCTTGATGGTGGAAAGCCCATGCAATGATCTC 2216
Qy 714 SerIleValAsnProAlaArgLysGlyLeuTyraGlnGlyThrGlyArgLeuPheGlu 733
Db 2217 TCTAAGCTCAATCTCGCCGCAAGAGTTATATAGAGGACAGAGAGGCTCTTCTTCAG 2276
Qy 734 PheTyraGlyLeuLeuHisAspAlaArgProLysGlnGlyAspAspArgProPheThr 753
Db 2277 TTTTACACCTGTGTGAATTTATACCCGCCCCAAGAGGGCGCAACCGTCATTTCTTGG 2336
Qy 754 LeuPheGluAsnValAlaIleMetGlyValSerAspLysArgAspIleSerArgPheLeu 773
Db 2337 ATGTTCCGAATGTTGTGGCCATGATGAAGTGAATGACAGAAAGACATTTCAAGATTCTG 2396
Qy 774 GluSerAsnProValMetIleAspAlaLysValSerAlaHisArgAlaArgTyra 793
Db 2397 GCAGTAAACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2456
Qy 794 PheTrpGlyAsnLeuProGlyMetAsnArgProLeuAlaSerThrValAsnAspLysLeu 813
Db 2457 TTCTGGGTTAACTTACCCGGAATGAACAGGCCGCTGATGATGATGATGATGATGATGAT 2516
Qy 814 GluLeuGlnGluCysLeuGluHisGlyValGlyIleAlaLysPheSerLysValArgThrIle 833
Db 2517 GAGCTGACGAGCTGCTGGAATTCATGATGACAGCAAGAAATTAAAGAAAGTGCAGCAATA 2576
Qy 834 ThrThrArgSerAsnSerIleLysGlnGlyLysAspGlnHisPheProValPheMetAsn 853
Db 2577 AACCACAACTGCAACTCATACAGACAGGCAAAAACAGCTTTCCCTGATGATGAT 2636
Qy 854 GluLysGluAspIleLeuTrpCysThrGluMetGluArgValPheGlyPheProValHis 873
Db 2637 GCGAAGAGCAGAGTTTGTGTGCTGATGAGCTGCAAAAGATCTTCCGCTTCCCTGCTAC 2696
Qy 874 TyrThrAspValSerAsnMetSerArgLeuValArgGlnArgLeuGluValArgSerTrp 893
Db 2697 TACACGAGACGTGTCAAATGAGCGCGCGCGCGCTGCAAGAGCTGCTGAGAGGCTCTGG 2756
Qy 894 SerValProValIleArgHisLeuPheAlaProLeuLysGluTyraPheAlaCys 911
Db 2757 AGTGAACGGTCACTACGACACTGTTTCCCTTGAAGACACTATTGCTGCT 2810

```

```

RESULT 8
AC009474 AC009474 168651 bp DNA linear PRI 08-NOV-2000
LOCUS

```

DEFINITION Homo sapiens BAC clone RP11-179G23 from 2, complete sequence.  
 ACCESSION AC009474  
 VERSION AC009474.4 GI:8749849  
 KEYWORDS HTG.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Sulerston, J.E. and Waterston, R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99083792  
 PUBMED 9847074  
 REFERENCE 1 (bases 1 to 168651)  
 AUTHORS Mulvane, E., Stoneking, T., LaPlante, Y. and McDill, B.  
 TITLE The sequence of Homo sapiens BAC clone RP11-179G23  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 168651)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-AUG-1999) Genome Sequencing Center, Washington  
 REFERENCE 4 (bases 1 to 168651)  
 AUTHORS University School of Medicine, 444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 TITLE Direct Submission  
 JOURNAL Submitted (27-JUN-2000) Genome Sequencing Center, Washington  
 REFERENCE 5 (bases 1 to 168651)  
 AUTHORS University School of Medicine, 444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 TITLE Direct Submission  
 JOURNAL Submitted (08-NOV-2000) Department of Genetics, Washington  
 University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT On Jun 27, 2000 this sequence version replaced GI:7630880.  
 -----  
 Center: Washington University Genome Sequencing Center  
 Web site: http://genome.wustl.edu/gsc  
 Contact: sapiens@wustl.wustl.edu  
 -----  
 Summary Statistics  
 -----  
 Center project name: H\_NH0179G23  
 -----  
 NOTICE: This sequence may not represent the entire insert of this  
 clone. It may be shorter because we only sequence overlapping  
 clone sections once, or longer because we provide a small overlap  
 between neighboring data submissions.  
 This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate  
 chemistry, or covered by high quality data (i.e. phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by sequence  
 from more than one subclone; and the assembly was confirmed by  
 restriction digest.  
 MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. John D.  
 McPherson, Department of Genetics, Washington University, St. Louis  
 MO. For additional information about the map position of this  
 sequence, see <http://genome.wustl.edu/gsc>  
 SOURCE INFORMATION:  
 The RP11-11 human BAC library was made from the blood of one male  
 donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,  
 Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved  
 approach for construction of bacterial artificial chromosome  
 libraries. Genomics 51:1-8. The clone may be obtained either from  
 Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong  
 and coworkers at the Roswell Park Cancer Institute  
 (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6  
 NEIGHBORING SEQUENCE INFORMATION:  
 The clone sequenced to the left is RP11-547F18; the clone sequenced  
 to the right is RP11-444B4, 200 base pair overlap. Actual end of  
 this clone is at base position 18711 of RP11-444B4.  
 Location/Qualifiers  
 1. 168651  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2"  
 /clone="RP11-179G23"  
 /clone\_lib="RPC1-11"  
 13...497  
 /rpt\_family="L2"  
 repeat\_region  
 616...683  
 /rpt\_family="L2"  
 repeat\_region  
 1615...1815  
 /rpt\_family="MIR"  
 repeat\_region  
 2577...2748  
 /rpt\_family="MIR"  
 repeat\_region  
 2954...3144  
 /rpt\_family="MIR1\_type"  
 repeat\_region  
 3149...3465  
 /rpt\_family="L1"  
 repeat\_region  
 3510...3567  
 /rpt\_family="MIR1\_type"  
 repeat\_region  
 3569...3760  
 /rpt\_family="MIR1\_type"  
 repeat\_region  
 3827...4198  
 /rpt\_family="MIR1\_type"  
 repeat\_region  
 5556...6129  
 /rpt\_family="MIR1\_type"  
 repeat\_region  
 6205...6316  
 /rpt\_family="MIR1\_type"  
 repeat\_region  
 6428...6686  
 /rpt\_family="L1"  
 repeat\_region  
 7247...7437  
 /rpt\_family="MIR3"  
 repeat\_region  
 7819...7868  
 /rpt\_family="CR1"  
 repeat\_region  
 9304...9469  
 /rpt\_family="Alu"  
 repeat\_region  
 10317...10439  
 /rpt\_family="MIR"  
 repeat\_region  
 12219...12455  
 /rpt\_family="Alu"  
 repeat\_region  
 12467...13537  
 /rpt\_family="L1"  
 repeat\_region  
 13553...14453  
 /rpt\_family="L1"  
 repeat\_region  
 14853...15123  
 /rpt\_family="MIR2\_type"  
 repeat\_region  
 16209...16345  
 /rpt\_family="MIR2\_type"  
 misc\_feature  
 16209...16345  
 /note="similar to EST AA338793 (NID:G1991103)"  
 misc\_feature  
 16210...16345  
 /note="similar to EST AI123298 (NID:G3539064) ov48c08.x1"  
 misc\_feature  
 16210...16345  
 /note="similar to EST AL037015 (NID:G5928008)"  
 repeat\_region  
 16513...16676  
 /rpt\_family="MIR1\_type"  
 repeat\_region  
 16970...17198  
 /rpt\_family="MIR1\_type"  
 repeat\_region  
 17199...17548  
 /rpt\_family="MIR1\_type"  
 repeat\_region  
 17575...18261  
 /rpt\_family="L1"  
 repeat\_region  
 19409...19531  
 /rpt\_family="L1"  
 repeat\_region  
 19540...19659  
 /rpt\_family="Alu"  
 repeat\_region  
 25485...25822  
 /note="similar to EST AI928247 (NID:G5664211) wo95h11.x1"

```
repeat_region      26074..26366
                    /rpt_family="Alu"
repeat_region      26384..26542
                    /rpt_family="MIR"
repeat_region      26569..27058
                    /rpt_family="u2"
repeat_region      27140..27647
                    /rpt_family="u2"
repeat_region      30824..31150
                    /rpt_family="MALR"
repeat_region      32262..32525
                    /rpt_family="CR1"
repeat_region      32568..32770
                    /rpt_family="MIR"
repeat_region      32775..33032
                    /rpt_family="MIR"
repeat_region      33172..33418
                    /rpt_family="MIR"
repeat_region      34350..35230
                    /rpt_family="Alu"
repeat_region      35302..35555
                    /rpt_family="u2"
repeat_region      35975..36081
                    /rpt_family="u2"
repeat_region      36364..36515
                    /rpt_family="Alu"
repeat_region      37266..37340
                    /rpt_family="u2"
repeat_region      37400..37440
                    /rpt_family="u2"
repeat_region      37615..37715
                    /rpt_family="MIR"
repeat_region      38006..38177
                    /rpt_family="MIR"
repeat_region      43171..43204
                    /rpt_family="u2"
repeat_region      44777..44926
                    /rpt_family="MIR"
repeat_region      45905..46110
                    /rpt_family="MIR"
repeat_region      48479..48638
                    /rpt_family="MIR"
repeat_region      48914..49033
                    /rpt_family="MIR"
repeat_region      49069..49337
                    /rpt_family="MIR"
repeat_region      49340..49621
                    /rpt_family="MIR"
repeat_region      50377..50498
                    /rpt_family="MIR"
repeat_region      56393..56545
                    /rpt_family="MIR"
repeat_region      58109..58420
                    /rpt_family="MIR"
```

## Alignment Scores:

```
Pred. No.:      4.54e-79      Length:      168651
Score:          2227.50      Matches:      488
Percent Similarity: 68.35%      Conservative: 67
Best Local Similarity: 60.10%      Mismatches: 137
Query Match:    45.14%      Indels:      131
DB:            9              Gaps:          9
```

US-09-720-086-7 (1-912) x AC009474 (1-168651)

```
QY 210 Trrpysaragluualagluylalalyvalialaglymetasnaalavalgluclu 229
Db 74844 TGGCTCATGGCTGAGAGAAAGCCGAGGTCACTGCAATGAATGAATGCTATGAGAA 74903
QY 230 Aasnnglyprogljgluserglnlyvalagluualasertproalavalgluclu 249
Db 74904 AACNAGGGCTCAGAGGATCTGCAGAAAGTGAAGAGCCAGCCCTGCTGTGGGAG 74963
QY 250 ProThraspProvalasertproThrvalalathrThrProgluProvalaglyseraspa 269
```

```
Db 74964 CCCAGCAACCCCACTCCCACTGCTCTTACCAACCTGAGCTGTGGGCTAGGCT 75023
QY 270 GLysaplysaanaIatThrlyalaglyaspaspjupprogluTyrgluaspjlyargly 289
Db 75024 GGGACCAAG---TACATAAAGCAGGTACAGTAACAGATACCAAGTACAGCTGGGAG 75080
QY 290 PheglylleglygluleuvaltrpglylyseuargglyPheSetrTrpProglyarg 309
Db 75081 TTGGCAATGGGGAGCTGTGTGAGGAAACTGCAGGGCTTCTCTGTGGCCG--- 75135
QY 310 lIeValserTrpTrpMetThrGlyargserargalalagluTyhArgTrpValMet 329
Db 75136 ATTGTCTTGCTGTGATGATGACAGCCAAAGCCAGAGAGCTGAAGGACCACTGTGTATG 75195
QY 330 TrpPheglyaspjlylysePheSerValValysvalagluylsleuMetProleuSer 349
Db 75196 TGGCTTGAGA-GGTGAGTTCTCAGTCTGTGTGTAGGAAGCTGTATCCACTGAGCTCC 75254
QY 350 PheCySerAlaPheHieglalathr-----TyraenlysglnPrometTyraArglys 367
Db 75255 TTTCGACGTGTCTTCCACCAACAC-CTATATATCAACAAGCAGCCATGTACCAAA 75313
QY 368 AlalIeTyrgluvalleuGlnvalalaserSerargalaglylyseuPheProvalCys 387
Db 75314 GCCATCTACGAAGCTCTGCGCGTGCCAGCAGTGT-ACAGGGAGAGCTTCTCAGGAGC 75372
QY 388 HisaspseraspjlyseraspThrAlalyalavalagluValglnaslyPrometIle 407
Db 75373 CATGACGTATGAGAGTACAGACCTGCCAAGGATACAGAGGTGACAGCAAGATATTC 75432
QY 408 GluTrpAlaleuglyglyPheglInProserglyProlysglyleuGluProProglu 427
Db 75433 AGATGACCTTTGGGGGTTTCCAGCTTCTCCCAAGAGACTGGAGCCCAAGAAAG 75492
QY 428 GluIysaenProTyrglygluvalTyrrhAspmetTrpValgluProgluAlalala 447
Db 75493 GAGAAAGATCCCAACAAAGATTATGACACGTGAGGGGTGACCTGAGGCAACTCT 75552
QY 448 TyrrlaPProProProAlalylylePProarglyysertThralagluys----- 464
Db 75553 TAGGACACCAACCCCAACAGCCCAAGAGCCCGAAAGACACAGCTTAAGAACTCAACATG 75612
QY 464 ----- 464
Db 75613 GGACCAAGCAGTAGCTCACCTGTATCCAGCACCTTGGAGGCTGAGCGGCGG 75672
QY 464 ----- 464
Db 75673 ATCAAGAGTCAGAGATAGAGACTGTGACATGTCAGGCCAACTGGTGAACCCGCTT 75732
QY 465 -----Prolys----- 466
Db 75733 CTACTAAAGTACCAAAATAGCTGACATGTGTCATGTGCTGTATCCAGTACTG 75792
QY 466 ----- 466
Db 75793 GGGAGGCTGAGGCGAGAGATCACTTGAACCGAGAGTGGAGGTTACAGTGAAGCCGAGA 75852
QY 466 ----- 466
Db 75853 TAGCACCATTCACCTCAGCCTGTGACAGAGCAAGACTGTTCAAAAAATAAATAA 75912
QY 467 -----VallysgluIlelleaspGluArgThrArggluArgleuValTyrgl 482
Db 75913 AAAAAAGCTGCACATCAAGGATGTGTGAAGAAC-ACAGMAAGATAGTATGATACAA 75970
QY 482 uValArgGlnlyCyserArgaenIlegluAspIleCysIleSerCysglySerleuAaenVa 502
Db 75971 AGTGAAGAAGAGTCCAGAAATCAAGAGCTTGTGTCTTGTGTAGAGACTCAATGT 76030
QY 502 lThleuGluHisProleuPheValglylyMetCysGlnAsnCyslysaenCysPhele 522
```

Db 76031 CACTCTGGAAACAGCCCTCTTCTATGCTAGGATATGACGACATTCGAAAGAACTGCTTCCT 76090  
QY 522 uGluCysAlaTyrGlnTyrAspAspGlyTyrGlnSerTyrCysThrIleCysCysG1 542  
Db 76091 GGAATATGATACCATGATGATGAAGAGGCTATCAGTCTCTGTCACCATCTCTGCTG 76150  
QY 542 YGlyArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluC 562  
Db 76151 TGGCAGAGGTGCTCATGTGCTGG---AACAACTGCTGCAGTGCTTTTGGGTGGAGTG 76207  
QY 562 sValAspLeuLeuValGlyProGlyAlaIleGlnAlaIleLysGluAspProTyrAs 582  
Db 76208 TATAGACCTCTTGGTGGCCAGAGGTGGCCAGGTGACCATTAAGAAAGATCCCTGGAA 76267  
QY 582 nCysTyrMetCysGlyHisLysGlyThrTyrGlyLeuLeuArgArgGluAspTyrPr 602  
Db 76268 CTCTCATGTGTGAACAAAGGCGCATCTGTGAGCTGCTGCGAGTGGCGGATGACTGGCC 76327  
QY 602 oSerArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProTyrAs 622  
Db 76328 CTCTAGCTCCAGATGTTCTTCCCAATAACCTAACCCAGGAATTTGACCTCTCGGAAGT 76387  
QY 622 lTyrProValProAlaGluLysArgLys---ProIleArgValLeuSerLeuPheAs 641  
Db 76388 CTACTCCCTGCTCCAGCTGAGAAAGAGCGCCCATCCAGGTGCTGCTCTTTTGA 76447  
QY 641 pGlyIleAlaThrGlyLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrI1 661  
Db 76448 CGGATTATGATGAGCTCTCTGCTGCTGAATGCTTGGCGCATTTAGGTAGATCACTACAT 76507  
QY 661 eAlaSerGluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysI1 681  
Db 76508 GTC-TGAGAGTGTGTGAGAGCGCATCAITGTGGCGATGCTGTGGCACCAGGGAGAT 76566  
QY 681 eMetTyrValGlyAspValArgSerValThrGlnLysHisIleGlnGluTyrGlyProPh 701  
Db 76567 CATGCACATCAGGATACCCAGAACCATCATACAGAGCAACACAGAGTGGGCCATTT 76626  
QY 701 eAspLeuValIleGlyGlySerProCysAsnAspLeuSerIleValAsnProAlaArgL 721  
Db 76627 CTC-----TGCAATGACCTCTCTCATGTCAACCTCGTCCACGCA 76664  
QY 721 sGlyLeuTyrGluGlyThrGlyArgLeuPheGluPheTyrArgLeuLeuHisAspAl 741  
Db 76665 AGGACCTTCAAGGCGACTGGCCAGCAATTTCTTGAGTGTCTCACCTGTG-CATCATGC 76723  
QY 741 aArgProLysGluGlyAspAspArgProPhePheTyrLeuPheGluAsnValAlaLame 761  
Db 76724 ACAGCCAAAGGAGGAGATGATCA-CCCTTCTTGTGGCTCCTTGAGAGTGTGTGGCCAT 76782  
QY 761 tGlyValSerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAs 781  
Db 76783 GGGCACTAGTCAAAAGAGAGACATCTCACAATTTTTTATTTCGAATTCATGATGATTA 76842  
QY 781 pAlaLysGluValSerAlaIleHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMe 801  
Db 76843 TGGCAAGAGTGTGAGTGCACACAGACTACTACATCTGTGTACCTTCCAGTAA 76902  
QY 801 tAsnArgProLeuAlaSerThrValAsnAspLysLeuGluLeuGlnCysLeuGluHi 821  
Db 76903 GAGCAGGCTATGGCATTCATATGAATGT-AACTGGAGCTGCGAGAGTGTGTGGAGCA 76961  
QY 821 sGlyArgIleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleL 841  
Db 76962 TGGCAAGAAAGCAAGGTGACAAAGCGAGGACCATTAACAGGTCAAACTCCATTAAT 77021  
QY 841 sGlnGlyLysAspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTyrCy 861  
Db 77022 GCAGGGCAAGACACAGCTGCTCCCTGCTTCATGAATGAGAAG---GACACCTTAGG-TG 77077  
QY 861 sThrGluMetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSe 881  
Db 77078 CATTGAATTTGAAGGGTATTGGACCTCCCTGCTCCATTATTAATGATCTCCAAACATAAG 77137

QY 881 rArgLeuAlaArgGlnArgLeuLeuGlyArgSerTyrSerValProValIleArgHisLe 901  
Db 77138 CCATTGGGCAAGGCAGAGACTGTGCTGAGTCATGAGCATGAGCATGCTACCTTGCCTAAT 77197  
QY 901 uPheAlaProLeuLysGluTyrPheAlaCysVal 912  
Db 77198 CTTCTCTCCACTGAAGGGTATTTTCTTGTGTG 77231

RESULT 9  
AF068627 4135 bp mRNA linear ROD 06-DEC-1999  
LOCUS Mus musculus DNA cytosine-5 methyltransferase 3B2 (Dmmt3b) mRNA,  
DEFINITION alternatively spliced, complete cds.  
ACCESSION AF068627  
VERSION AF068627.2 GI:6449471  
KEYWORDS  
SOURCE Mus musculus.  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 4135)  
AUTHORS Okano,M., Xie,S. and Li,E.  
TITLE Cloning and characterization of a family of novel mammalian DNA  
(cytosine-5) methyltransferases  
JOURNAL Nat. Genet. 19 (3), 219-220 (1998)  
MEDLINE 98324766  
PUBMED 9662389  
REFERENCE 2 (bases 1 to 4135)  
AUTHORS Xie,S., Okano,M. and Li,E.  
TITLE Direct Submission  
JOURNAL Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA  
REFERENCE 3 (bases 1 to 4135)  
AUTHORS Okano,M., Chijiwa,T., Sasaki,H. and Li,E.  
TITLE Direct Submission  
JOURNAL Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA  
REMARK Sequence update by submitter  
COMMENT On Nov 19, 1999 this sequence version replaced gi:3327981.  
FEATURES  
source  
1. 4135  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/map="84.0 cm"  
1. 4135  
/genes="Dmmt3b"  
1. 4135  
/genes="Dmmt3b"  
/notes="similar to EST sequences deposited in GenBank  
Accession Numbers AA116694, AA119979, AA177277, AA210568,  
AA407106, and AA575617"  
269, 2788  
/genes="Dmmt3b"  
/function="de novo DNA methylation"  
/notes="alternatively spliced product; contains Cys-rich  
region; C-terminal region is similar to corresponding  
region of DNA cytosine-5 methyltransferase 3A"  
/codon\_start=1  
/product="DNA cytosine-5 methyltransferase 3B2"  
/protein\_id="AAC40179.2"  
/db\_xref="GI:6449472"  
translation="MKGDSRLHNEEGAGCYECIIIVGNFSDQSSDTKDAPSPVLE  
AICPEVCTPETRRSSRLSKREVSLLNYTQDMTGDGDRDDEVDGDSILMPK  
LTRYKDTKTRSESPAVRTHRSNGTSLSERASPRITRGRQGRHHVQVEYEPFATR  
SRRRSASSTPMSPASVDPMEEVTPKSVSTPSVDLSQDQEGMDTTQVDAESRD  
GDSLEYQDDKEFGIDLVWGIKGFSSWPMVVMKATSKQRMFGMRVQVDFGDKRP  
SETSDPKLVALGFSQHFNLATFNKLVSKYKMYHTLEKARVAGKTSSPFGSLSD  
QLKPMLEWAHGGFKPTGIEGLKPNKPKENKSRRTTNDASAESPPPKRLKNTSYGG  
KDRGEDSERMASVEITNKGKLEDCISCKGNKPVSFHPLFEGGLCQSCRDLFLEL  
FMYDDEGYSYCTVCCCEGRELSCNTSCRCFCVECLEVLVGATGAEDAKLQEPWS  
CYMCLPQRCHVLRRRKQWNRMLQDFTTDPDLEEPPEPKLYPAIPAAKRRPRLVLSL

FDGIATGYLVKELGIKVEKYIASVCAESIAVGVKHEGOIKYVNDVKRTKNTIE  
WGPFDLVTIGSSPCNDLSNVNPAKGLVEGRLPFERYHLANTRPREGNRPFWF  
ENNVAMKNDKDISRPACNPVMDIKVSAHARAYFWNLPGMRPYMAKNDKL  
ELDDCLEFSTAKLKVQITTTSSKNSIROKNOLEPVMVNGDVLMTCLERIIFGP  
AHTDYSNMGRKROKILGRSWSVPIRHLFAPLKYFACE"

BASE COUNT 1043 a 1048 c 1106 g 938 t  
ORIGIN

## Alignment Scores:

Pred. No.:	1,62e-80	Length:	4135
Score:	2226.00	Matches:	455
Percent Similarity:	59.71%	Conservative:	129
Best Local Similarity:	46.52%	Mismatches:	254
Query Match:	45.11%	Indels:	140
DB:	10	Gaps:	15

US 720-086-7 (1-912) x AF068627 (1-4135)

Qy 16 AAlaAgluaArgluGluAAspArgLysApqLysgluGluGluGluProAlaArgLys 35  
| | | | |  
| | | | |  
Db 23 GCGGCCCAAGTAAAGTACGACGATCGGCGCGGAGATTCCGAAACC-----GAC 76  
Qy 36 GluGluAArgluGluProSerThrAlaArgLysValGlyArgProGlyArgLysArg 55  
| | | | |  
| | | | |  
Db 77 ACTCCGCGCCCGCGCGGACCGCGCGGATCGGCGCGCGCTACAGCCAG 136  
Qy 56 LysHisPro---ProValGluSerGlyAspThrProLysAspProAlaValIle----- 72  
| | | | |  
| | | | |  
Db 137 CCTCAGCAGAGCCCGCTGAGGCTTGCCAGACCTTGAAACCTCAGGTTATACCTTT 196  
Qy 73 -----SerLysSerProSerMet 78  
| | | | |  
| | | | |  
Db 197 CCAGACGGGATCTCCCTCCCATCCATAGTCCTTGAGACCAATCCAGGGCTTC 256  
Qy 79 AAlaGluAAsp-----SerLys 83  
| | | | |  
| | | | |  
Db 257 TTTCCGGAAGCAATGAGAGACACAGACATCTGAATGAAGAGAGTCCCGCGCG 316  
Qy 84 AAlaSerGlu---LeuLeuProAsnGlyAspLeuGluLysArgSerGluProGluProGlu 102  
| | | | |  
| | | | |  
Db 317 TATGAGAGTGATATATGTTATATGGAATTCACGTACCGCTCTCAGACACGAGAGAT 376  
Qy 103 GluGluSerProAlaGlyGlyGluLysGlyGlyAlaProAlaGluGlyGlyGlyAla 122  
| | | | |  
| | | | |  
Db 377 GCTCCCTCACCC----- 388  
Qy 123 GluThrLeuProGluAlaSerArgAlaValGluAsnGlyCysCysThrProLys----- 140  
| | | | |  
| | | | |  
Db 389 -----CCAGTCTTGAGGCAATCTGCACAGACCAAGTCTGCACACAGAGACAGA 439  
Qy 140 ----- 140  
Db 440 GGCCGCAAGTCAAGCTCCCGGCTGTCTAAGAGGAGGTCTCCAGCTTCTGAATTACAG 499  
Qy 141 -----GluGlyArgGlyAlaPro 146  
| | | | |  
| | | | |  
Db 500 CAGGACATGACAGAGATGAGACAGAGATGATGAAGTAGATGAGATGCGCTGTGAT 559  
Qy 147 AAlaGluAgluAgluLysGluGluLysGluThrAsnIleGluSerMetLysMetGlySer 166  
| | | | |  
| | | | |  
Db 560 ATTCTTAATGCCAAAGCTCACCGGTGAGACCAAGACACCGAGACCGCTCTAAGACCG 619  
Qy 167 ArgGlyArgLeuArgGlyGlyLeuGlyLysGlyLysSerLeuArgGlyArgProMetPro 186  
| | | | |  
| | | | |  
Db 620 GCTGTCCGAACCGCATAGCAATGGGACCTCAGCTTGAGAGCGAAGACCTTCCCC 679  
Qy 187 ArgLeuThr-----PheGluAlaGlyAspPro 195  
| | | | |  
| | | | |  
Db 680 AGAATCACCCAGAGTCCGAGAGGCCGCCACATGTCAGAGATACCTCTGAGATTTCGG 739  
Qy 196 TyrTyrLysSerLysArgLysArgAspGluTyrPleuAlaArgTyrPlyArgGluAlaGlu 215  
| | | | |  
| | | | |  
Db 740 GCTACAGAGTCTCGAGAGAGTCA----- 763

Qy 216 LysLysAlaLysValIleAlaGlyMetAsnAlaValGluGluAsnGluGlyGlu 235  
| | | | |  
| | | | |  
Db 764 -----GCATGCTCTTCAGCAAGACCGCATGATCATCCCGCCAGC 805  
Qy 236 SerGluLysValGluGluAlaSerProProAlaValGluGluProThrAspProAlaSer 255  
| | | | |  
| | | | |  
Db 806 GTGCACTTCATGAAAGTACAGACCTTAAGAGCGCTGATACCA-----TCA 853  
Qy 256 ProThrValAlaThrThrProGluProValGlySerAspAlaGlyAspLysAsnAlaThr 275  
| | | | |  
| | | | |  
Db 854 GTTGACTTGAGCCAGAGATGAGATGAGAGGTATGATATCCACACAGGTGATGCGAG 913  
Qy 276 LysAlaGlyAspAspGluProGluTyrGluAspGlyArgGlyPheGlyIleGlyGluLeu 295  
| | | | |  
| | | | |  
Db 914 AGCAGAGATGAGACAGACAGAGTATCAGATGATGAAGATTGGAATAGGTGACCTC 973  
Qy 296 ValTyrGlyLysLeuArgGlyPheSerTyrTyrProGlyArgIleValSerTyrProMet 315  
| | | | |  
| | | | |  
Db 974 GTGAGGGAAGATCAAGGCGCTCTCTGTGGCTCCATGTGGTGTCTCGAAGCC 1033  
Qy 316 ThrGlyArgSerArgAlaAlaGluGlyThrArgTyrValMetTyrPheGlyAspGlyLys 335  
| | | | |  
| | | | |  
Db 1034 ACTTCCAGAGACAGGCCATGCCGGAATGCCGTGGTATCAGTGCTTGGTATGCGAAG 1093  
Qy 336 PheSerValValCysValGluLysLeuMetProLeuSerSerPheCysSerAlaPheHis 355  
| | | | |  
| | | | |  
Db 1094 TTTTCTGAGATCTCTGCTGACAAACCTGTGGCTGTGGGCTGTTCACGACATTATAT 1153  
Qy 356 GluAlaThrTyrAsnLysGluProMetTyrArgLysAlaIleTyrGluValLeuGluVal 375  
| | | | |  
| | | | |  
Db 1154 CTGCTACTACTTCAATTAAGCTGTGTTCTTATAGAAAGCCATGTAACACTGTGAAGAAA 1213  
Qy 376 AAlaSerSerArgAlaGlyLysLeuPheProValCysHisAspSerAspGluSerAspThr 395  
| | | | |  
| | | | |  
Db 1214 GCCAGGCTTCAGCTGCGAAGACCTTC-----TCCAGAGTCT 1252  
Qy 396 AAlaLysAlaValGluValGluAsnLysProMetIleGluTyrAlaLeuGlyGlyPheGlu 415  
| | | | |  
| | | | |  
Db 1253 GGAAGATCAGTCGAGAGACAGCTGAGCCATGCTGAGTGCGCCACAGGTGCTTCAG 1312  
Qy 416 ProSerGlyProLysGlyLeuGluProProGluGluGlu---LysAsnProTyrLysGlu 434  
| | | | |  
| | | | |  
Db 1313 CCTACTGGGATCGAGGCGCTCAAAACCAACAAAGAAACCAAGAAACCAAAAGTCCAGA 1372  
Qy 435 ValTyrThrAspMetTyrValGluProGluAlaAlaIleTyrAlaProProProAla 454  
| | | | |  
| | | | |  
Db 1373 CGCACACCAAT-----GACTCTGCTGCTTGTGATCCCCCACCACAG 1417  
Qy 455 LysLysProArgLysSerThrAlaGluLysProLysValLysGluIleIleAspGluArg 474  
| | | | |  
| | | | |  
Db 1418 CGCCTCAAGAAATATAGTATGAGGCGGAAGACCGAGGGAG-----GATGAGAG 1468  
Qy 475 ThrArgGluArgLeuValTyrGluValArgGlyLysCysArgAsnIleGluAspLysCys 494  
| | | | |  
| | | | |  
Db 1469 AGCCGAGAACGATGAGCTTCTGAAGTCCACCAACAAAGGCAATCTGGAAGACCGCTGT 1528  
Qy 495 IleSerCysGlySerLeuAsnValThrLeuGluIleIleProLeuPheValGlyMetCys 514  
| | | | |  
| | | | |  
Db 1529 TTGCTCTGTGAAAGAAAGAACCTGTGTCTTCAACCCCTCTTGAAGGGGTCTGT 1588  
Qy 515 GluAsnCysLysAsnCysPheLeuGluCysAlaTyrGluTyrAspAspAspGlyTyrGlu 534  
| | | | |  
| | | | |  
Db 1589 CAGAGTCCCGGATCGCTTCTGAGAGCTCTTCAATGATGATGAGAGAGGCTATAG 1648  
Qy 535 SerTyrCysThrIleCysCysGlyGlyArgGluValLeuMetCysGlyAsnAsnAsnCys 554  
| | | | |  
| | | | |  
Db 1649 TCTTACTGACACGATGCTGTGAGGCGGTGAACCTGCGCTGTACAGTACACACAGCTGC 1708  
Qy 555 CysArgCysPheCysValGluCysValAspLeuValGlyProGlyAlaAlaGluAla 574  
| | | | |  
| | | | |  
Db 1709 TGCAGATGCTTCTGTGTGAGATGTGTGAGGTGCTGTGAGGCGGACAGCAGCTGAGAT 1768





QY 38 gGlnGluProSerThrThrAlaArgLysValGlyAArgProGluYArgLysHisPr 58  
 Db 228 CCGCCCCCGCCAGAGACCCGCGCGCATCGCGCGCTACAGCCAGCCCTCACGA 287  
 QY 58 o---ProValGluSerGlyAspThrProLysAspProAlaValIle----- 72  
 Db 288 CAGGCCCGCTAGGCTTGTGCCAGACCTTGGAAACCTCAGATATACCTTTCCAGAGCC 347  
 QY 73 -----SerLysSerProSerMetAlaGlnAs 81  
 Db 348 GCGATCTCCCTCCCGCCATCATAGTGCCTTGGACCAATCCAGGCGCTTCTTCAAGA 407  
 QY 81 p-----SerGlyAlaSerG 86  
 Db 408 AACATGAAGGAGACACAGACATCTGATGAAAGAAAGAGGAGGTCGACGGGTATGAGGA 467  
 QY 86 u---LeuLeuProAsnGlyAspLeuGluLysArgSerGluProGluGluGlySe 105  
 Db 468 GTGCATTATCGTTAATGGAACTTCATGACCGACGCTCTCAGACAGAGATGCTCCCTC 527  
 QY 105 rProAlaGlyGlyGluLysGlyAlaProAlaGluGluGlyAlaAlaGluThrIle 125  
 Db 528 ACCC----- 531  
 QY 125 uProGluAlaSerAlaValAlaGluAsnGlyCysCysThrProLys----- 140  
 Db 532 -CCAGTCTTGAGAGCAATCTGCACAGAGCCAGTCTGCACACAGAGACAGAGCCGAG 590  
 QY 140 ----- 140  
 Db 591 GTCAAGCTCCCGCTGTTAAGAGGAGGTCTCCAGCTTCTGATTAACAGCAGAGACAT 650  
 QY 141 -----GluGlyArgGlyAlaProAlaGluAla 149  
 Db 651 GACAGAGATGAGACAGACAGATGATGATGATGATGATGATGATGATGATGATGAT 710  
 QY 149 agLyLysGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 169  
 Db 711 GCCAAGCTCAGCCGTGAGACCAAGAGACACAGAGCGCGCTGAAACCCGCGCTGCG 770  
 QY 169 gLeuArgGlyGlyLeuGlyTyrGluSerSerLeuArgGluGluGluGluGluGluGlu 189  
 Db 771 AACCCGATATGACATGGGACCTCCAGCTTGAGAGGACCAAGAGCCCTCCAGATAC 830  
 QY 189 r-----PheGlnAlaGlyAspProTyrTyrT 198  
 Db 831 CCGAGGTGGCAGGCGCCACCATGTCAGAGATACCTTGAGATTTCGGGCTACAG 890  
 QY 198 eSerLysArgLysArgAspGluTyrLeuAlaArgTyrLysArgLysAlaGluLysAla 218  
 Db 891 GTCGCGAGAGCTGCA----- 906  
 QY 218 aLysValIleAlaGlyMetAsnAlaValGluGluGluGluGluGluGluGluGluGlu 238  
 Db 907 -----GCATCTGCTCTGCAAGACCGCATGTCATCCCTCCAGAGGTGACTT 956  
 QY 238 aValGluGluLysArgProProAlaValAlaGluGluGluGluGluGluGluGluGluGlu 258  
 Db 957 CATGGAAGAGTACACCTTAAGCGGTGATACCA-----TCAATTGACTT 1004  
 QY 258 lAlaThrThrProGluProValGlySerAspAlaGlyLysAspLysAsnAlaThrLysAla 278  
 Db 1005 GAGCAGAGATGAGATGATGAGAGGATGATGATGATGATGATGATGATGATGATGAT 1064  
 QY 278 yAspAspGluProGluTyrGluLysGlyArgGlyPheGlyTyrGlyGluLeuValTyrG 298  
 Db 1065 TCGAGACGACGACAGAGTATCAGATGATGATGATGATGATGATGATGATGATGAT 1124  
 QY 298 yLysLeuArgGlyPheSerTyrProGluYArgLysValSerTyrProMetThrGlyArg 318  
 Db 1125 AAAGATCAAGGGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1184

QY 318 gSerArgAlaAlaGluGlyThrArgTyrValMetTyrPheGlyLysArgLysPheSerVal 338  
 Db 1185 GCGACAGGCCATGCCCCGGAATGCCCTGGTACAGTGTGTGTGTGTGTGTGTGTGTGTGA 1244  
 QY 338 lValCysValGluLysLeuMetProLysSerSerPheCysSerAlaPheHisGlnAlaThr 358  
 Db 1245 GATCTCTGCTACAACTGGTGGCTCTGGGCTTTACAGCCAGCATTTATCTGGCTAC 1304  
 QY 358 rTyrAsnLysGluProMetTyrArgLysAlaIleTyrGluValLeuGluValAlaSerSe 378  
 Db 1305 CTTCATATAGCTGTTCTTATAGAAAGCCATGATACACACTTGGAGAAAGCAGAGGT 1364  
 QY 378 rArgAlaGlyLysLeuPheProValCysHisAspSerAspGluSerAspThrAlaLysAl 398  
 Db 1365 TCGAGCTGGCAAGCCTTC-----TCCAGCAGTCTCGAGAGTCC 1403  
 QY 398 aValGluValGluLysProMetIleGluTyrAlaLeuGluGlyPheGlnProSerG 418  
 Db 1404 ACTGAGAGACGAGCTGAAGCCCATGCTGAGTGGGCCACAGCTTCAAGCTTACTGG 1463  
 QY 418 yProLysGlyLeuGluProProGluGluGluLysAsnProTyrLysGluValTyrTh 437  
 Db 1464 GATGAGAGGCTCTAAACCAACCAAGAACCAAGAACCAAGAACCAAGAACCAAGAAC 1523  
 QY 437 rAspMetTyrValGluProGluAlaAlaAlaTyrAlaProProProProAlaLysLysPr 457  
 Db 1524 CAAT-----GACTCTGCTGCTTCTGAGTCCCTCCACCAAGCCCTCA 1568  
 QY 457 oArgLysSerThrAlaGluLysProLysValLysGluIleLeuSerGluGluThrArg 477  
 Db 1569 GACAAATGCTATGCGGAGGAGACGAGGGAG-----GATAGAGAGACCGAGGA 1619  
 QY 477 uArgLeuValTyrGluValArgGluLysCysArgAsnIleGluAspIleCysIleSerCy 497  
 Db 1620 ACGAGATGCTTCTGAATCAACCAACCAAGAGCAATCTGGAACCCCTTTGTCTGTG 1679  
 QY 497 sGlySerLeuAsnValThrLeuGluHisProLeuPheValGlyGlyMetCysGlnAsnCy 517  
 Db 1680 TGGAAAGAGAACCTGCTGCTCCACCCCTCTTGAAGGAGGAGGCTCTGCAAGATTG 1739  
 QY 517 sLysAsnCyPheLeuGluCysAlaIleTyrGluTyrAspAspAspGlyTyrGlnSerTyrCy 537  
 Db 1740 CCGGATCGCTTCTTAAGCTCTTCTACATGATGATGATGATGATGATGATGATGATG 1799  
 QY 537 sThrIleCysCysGlyGlyAlaArgGluValLeuMetCysGlyLysAsnAsnAsnCySAspCy 557  
 Db 1800 CACCTGTGCTGATGAGGCGCTGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1859  
 QY 557 sPheCysValGluCysValAspLeuLeuValGlyProGlyValAlaAlaAlaIleLys 577  
 Db 1860 CTTCGTGTGAGAGTCTGAGAGGCTGAGGCTGAGGCGCAGGCACTGAGATGCAAGCT 1919  
 QY 577 sGluAspProTyrPheCysTyrMetCysGlyHisLysGlyThrTyrGlyLeuLeuArg 597  
 Db 1920 GCAGGAACCCCTGAGCTGCTATGCTGCTCCCTCAGGCGCTGATGGGGCTCTCCGAG 1979  
 QY 597 gArgGluAspTyrProSerArgLeuGluMetPhePheAlaAsnAsnHisAsp---GlnG 616  
 Db 1980 CAGGAAGATTGGAACATGCGCTGCAAGACTTCTTCACTAATCTGCTGCTGCAAG 2039  
 QY 616 uPheAspProProLysValTyrProProValProAlaGluLysArgLysProIleArgVa 636  
 Db 2040 ATTTGAGACCAACCAAGTTGTATCCAGCAATTCCTGCAAGCCAAAGAGAGGCCATTAGAGT 2099  
 QY 636 lLeuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeuLysAspLeuGlyIleG 656  
 Db 2100 CCTGCTCTGTTGATGGAATGCAAGGGGTACTTGTGTCTCAAGAGAGTTGGGTATTA 2159  
 QY 656 nValAspArgTyrIleAlaSerGluValCysGluAspSerIleThrValGlyMetValAr 676  
 Db 2160 AGTGAAAGAGTACATGCTCCGAAAGTCTGCAAGATCCATCCCTGAGGAACTGTAA 2219  
 QY 676 gHisGlnGlyLysIleMetTyrValGlyAspValArgSerValThrGluLysHisIleG 696

Db 2220 GCATGAAGCCAGATCAAAATATGTCAATGACGTCGCGAAGAAATCAACCAAGAAAAATATTGA 2279  
Qy 696 nGluTTPGlyProPheAspLeuValIleGlyGlySerProCysAsnAspLeuSerIleVa 716  
Db 2280 AGAGTGGGCGCCCTTGGCTGATGTTGGTGAAGCCATGCAATGATCTCTCTAAGCT 2339  
Qy 716 lAenProAlaArgLysGlyLeuTyTGlyGlyThrGlyArgLeuPheGluPheTyAr 736  
Db 2340 CAATCTGCGCCGCAAGGTTATATGAGGGCACAGGAAGCTCTTCTCGAGTTTACCA 2399  
Qy 736 gLeuLeuHisAspAlaArgProLysGluGlyAspArgProPheTTPLeuPheG1 756  
Db 2400 CTTGCTGAATTTATACCCGCCCAAGAGGCGCACACCTCATCTTCTCGAGTTTCA 2459  
Qy 756 uAenValValAlaMetGlyValSerAspLysArgAspIleSerArgPheLeuGluSerAs 776  
Db 2460 GAATGTTGTGGCCATGAAGTGAATGACAAAGAGACATCTCAAGATTCCTGGCATGTAA 2519  
Qy 776 nProValMetIleAspAlaLysGluValSerAlaAlaHisArgAlaArgTyPheTTPG1 796  
Db 2520 CCAGTGATGATCGATGCGATCAAGGTGTCTGCTGCTCACAGGGCCCGGTACTTCTGGG 2579  
Qy 796 YAsnLeuProGlyMetAsnArgProLeuAlaSerThrValAsnAspLysLeuGluLeuG1 816  
Db 2580 TAACCTACCGGAATGAACAGCCGCGTGATGGCTTCAAGAATGATAAGCTCGAGCTGCA 2639  
Qy 816 nGluCysLeuGluHisGlyArgIleAlaLysPheSerLysValArgThrIleThrThrAr 836  
Db 2640 GGACTCGCTGGAGTTCAGTAGGACAGCAAAAGTTAAAGAAAGTGCAGACAAATAACCCAA 2699  
Qy 836 gSerAsnSerIleLysGlnGlyLysAspGlnHisPheProValPheMetAsnGluLysG1 856  
Db 2700 GTCGAATCTCATCAGACAGGCGCAAAACACAGCTTTTCCCTGTAGTCATGATGCCAAGA 2759  
Qy 856 uAspIleLeuTTPCysThrGluMetGluArgValPheGlyPheProValHisTyThrAs 876  
Db 2760 CGACGTTTGTGTGACTGAGCTGCAAGAGATCTTCGGCTTCCCTGCTACTACACGA 2819  
Qy 876 pValSerAsnMetSerArgLeuAlaArgGlnArgLeuLeuGlyArgSerTTPSerValPr 896  
Db 2820 CGTGTCAACATGGCCCGCGCCGCTGAGAGCTGCTGGGCGAGTCTCTGGAGTGACC 2879  
Qy 896 oValIleArgHisLeuPheAlaProLeuLysGluTyPheAlaCys 911  
Db 2880 GGTCAATCAGACACCTGTTTGGCCCTTGAAGGACTACTTTGCTGT 2925  
RESULT 11  
AF151973  
LOCUS  
DEFINITION  
Mus musculus DNA cytosine-specific methyltransferase isoform 5  
ACCESSION  
AF151973  
VERSION  
AF151973.1  
KEYWORDS  
GI:8347127  
SOURCE  
Mus musculus.  
ORGANISM  
Mus musculus.  
REFERENCE  
1. (bases 1 to 4223)  
AUTHORS  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Yin.B., Chen.Y.T., Zhu.M., Luo.Y.J., Zhu.N., Xu.S.C., Wu.G.Y. and  
Shen.Y.  
TITLE  
Cloning of full-length Dnmt3b cDNA and its alternative splicing  
isoforms in mouse embryonic tissue  
JOURNAL  
Unpublished  
REFERENCE  
2. (bases 1 to 4223)  
AUTHORS  
Yin.B., Chen.Y.T., Zhu.M., Luo.Y.J., Zhu.N., Xu.S.C., Wu.G.Y. and  
Shen.Y.  
TITLE  
Direct Submission  
Submitted (17-MAY-1999) Department of Biochemistry and Molecular  
Biology, Institution of Basic Medical Sciences, Chinese Academy of  
Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,  
P.R.China

FEATURES  
source  
Location/Qualifiers  
1. 4223  
/organism="Mus musculus"  
/strain="KM"  
/db\_xref="taxon:10090"  
/dev\_stage="8-9 day old embryo"  
1. 4223  
/gene="Dnmt3b"  
297..2876  
/gene="Dnmt3b"  
/function="methylates cytosine in DNA"  
/note="alternatively spliced"  
/codon\_start=1  
/product="DNA cytosine-specific methyltransferase isoform 5"  
/protein\_id="AAF74519.1"  
/db\_xref="GI:8347128"  
translation="MKGSRSLNKEEGASGYECIIIVNGNFSDQSDSTKDAPSPPVV  
AICTEPCVTPETRRSSRLNKEEGASGYECIIIVNGNFSDQSDSTKDAPSPPVV  
LTHETKDTKTRSESAPVTRHNSGTSLERQASPRI TGRQGRHHVQVEYVEVEPATR  
SRRRASSTASPSSPASVDMEVTPKSVTPS VLSQDQEGMDTTQVDAESIY  
GDSEYQDDKEFGIDLVWGKIGFSWPMVAMVHTLEKARVAKGTFSFGSGLED  
SEISADKLVALGFSQHPNLATFNKLVSRKAMVHTLEKARVAKGTFSFGSGLED  
QLKPMLEWAHGKFTGIEGLKNKQPVVNSKVRSDSNLEPRRENKSRRTTN  
DSAASEPPKRLKTNISYGGKDEDESESRMASEVTNNKGNLCLSCCKENPVS  
FHP1FEGGLQSCDRFLFELFYMYDEGYQSYCTVCEGRELKLLCSNCCRCFVEP  
LELVGAGTAEDAKLQEPWCMCLPQRCGVLRKRWMLRQLQFFTTDPLEFEFP  
PKLYPAIPAAKRRPIRVLSLFDGIATGLYLKELGIVKVKYIASEVCAESIAGTVFV  
EGQIKVNDVRKI TKNIIEEMGPFDLVIGGSPCNLDSNVNPAKGLYEGTGRLEAFVY  
HLNLYTRPKEGDNRPFFWMPENVMKVKNDKDI SRFLACNPVMDAIKVGAHARAY  
FWGNLPGMNRPMVASKNDKLEQDLEFSRTAKLKVQITTTKNSIRQGNQLPPVV  
MNGKDDLVWCTELERIFGFPAPHTVDSNMGRGAROKLGRSWSVPIRHLFAPLKYF  
ACE"

BASE COUNT 1051 a 1084 c 1147 g 941 t

ORIGIN  
Alignment Scores:  
Pred. No.: 2,07e-80 Length: 4223  
Score: 223.50 Matches: 450  
Percent Similarity: 59.81% Conservative: 123  
Best Local Similarity: 46.97% Mismatches: 240  
Query Match: 45.06% Indels: 145  
DB: 10 Gaps: 14

US-09-720-086-7 (1-912) x AF151973 (1-4223)  
Qy 41 ProSerThrThrAlaArgLysValGlyArgProGlyArgLysArgLysHisProProVal 60  
Db 171 CCAAGTAAACGTAGCGAGCGATCGCGCGGAGATTCCGGAACCCGACACTCCGCGCG 230  
Qy 61 GluSerGlyAspThrProLys-----AspProAlaValIleSer 73  
Db 231 CCGCGCGCCAGGACCGCGCGCGATCGCGCGCGCGCTACAGCCAGCC-----TCA 284  
Qy 74 LysSerProSerMetAlaGlnAspSer-----GlyAlaSerGlu 86  
Db 285 CCACAGGAAACATGAGGAGGAGACAGACACATCTGATGAGAGAGAGGTCGACGCGG 344  
Qy 87 -----LeuLeuProAsnGlyAspLeuGluLysArgSerGluProGlnProGlu 102  
Db 345 TATGAGGAGTGCATTCGTTATGGAACTTCAGTGACCACTCTCAGACAGGAAGAT 404  
Qy 103 GluGlySerProAlaGlyGlyGlnLysGlyAlaProAlaGluGlyGluGlyAlaAla 122  
Db 405 GCTCCTCACC----- 416  
Qy 123 GluThrLeuProGluAlaSerArgAlaValGluAsnGlyCysCysThrProLys----- 140  
Db 417 -----CCAGTCTTTGGAGGCAATCTGCACAGAGCCAGTCTGCACACAGACCA 467  
Qy 140 ----- 140  
Db 468 GGCGCGAGGTCAAGCTCCCGGCTGTCTTAAGAGGAGGAGGTCTCCAGCTTCTGAATTACAG 527

Oy 141 ----- GluGlyArgGlyAlaPro 146  
 Db 528 CAGGACATGACAGAGATGGAGACAGAGATGATGAAGTAAAGTAAAGGAAATGGCTCGAT 587  
 Oy 147 AlaGluAlaGlyIysGluIysGluTyrAsnIleGluSerMetIysMetGluGlySer 166  
 Db 588 ATTGTAATGCCAAGAGCTCACCCGTGAGACCAAGACACACAGAGCCGCTCTGAAAGCCCG 647  
 Oy 167 ArgGlyArgLeuArgGlyGlyLeuGlyTyrGluSerSerLeuArgGluArgProMetPro 186  
 Db 648 GCTGTCCGAACCCGACACTACCAATGGGAGCTCCAGTGTGAAAGGCCAAGAGCCCTCCCC 707  
 Oy 187 ArgLeuThr-----PheGluAlaGlyAspPro 195  
 Db 708 AGAATTCACCCGAGGTCCGAGCGCCGCCACATGTGCAGAGACTACCTGTGAGATTCCG 767  
 Oy 196 TyrTyrIleSerIysAspGlySerAspArgIuTyrPheAlaArgTyrPheArgGluAlaGlu 215  
 Db 768 GCTACCAAGCTCTCGAACACTCA----- 791  
 Oy 216 LysIysAlaIysValIleAlaGlyMetCAsnAlaValGluGluAsnGluIProGlyGlu 235  
 Db 792 -----GATGTCCTTCAGCAAGACGCCATGTCATCCCTCCACGC 833  
 Oy 236 SerGluIysValGluGluAlaSerProProAlaValGluGluInProThrAspProAlaSer 255  
 Db 834 GTCGACTTCATGGAAGAGTACACCTTAAGAGGTCATGACCA-----TCA 881  
 Oy 256 ProThrValAlaThrThrProGluProValGlySerAspAlaGlyAspIysAsnAlaThr 275  
 Db 882 GTTGACTTGAGCCAGATGAGATCAGAGAGGTATGATGATACACACAGGTGGATGCAAG 941  
 Oy 276 LysAlaGlyAspAspArgIuProGluTyrGluAspGlyArgGlyPheGlyIleGlyGluLeu 295  
 Db 942 AGCATATATGGAGACAGACAGAGATCATGATGATTAAGATTGGAATAGTGACCTC 1001  
 Oy 296 ValTyrGlyIysLeuArgGlyPheSerTyrTyrProGlyArgIleValSerTyrTyrMet 315  
 Db 1002 GTGGGGGAAGAAGATCAAGGGCTTCTCGGTGGCCATGATGGTGTGTCGTAAGAACCC 1061  
 Oy 316 ThrGlyArgSerArgAlaAlaGluGlyThrArgTyrValMetTyrPheGlyAspGlyIys 335  
 Db 1062 ACCCTCAAGCACAGCGCCATGCCCGGAATGCCGTGGGTACATGGTTGGTGAATGGCAAG 1121  
 Oy 336 PheSerValValCysValGlyIysLeuMetProLeuSerSerPheCysSerAlaPheHis 355  
 Db 1122 TTTTCTAGAGCTCTGCTGCAACAATCGTGCGCTCGGGCTGTTCAGCCAGCATTTAAT 1181  
 Oy 356 GluAlaThrTyrAsnIysGluInProMetTyrArgIysAlaIleTyrGluValLeuGluVal 375  
 Db 1182 CTGGCTACCTTCAATTAAGCTGGTTCTTATAGGAAGGCCATGTACACACTGTGAGAA 1241  
 Oy 376 AlaSerSerArgAlaGlyIysLeuPheProValCysHisAspSerAspIuSerAspThr 395  
 Db 1242 GCCAGGGTTGAGCTGCAAGACCTC-----TCAAGCAGTCTT 1280  
 Oy 396 AlaIysAlaValGluValGluAsnIysProMetIleGluTyrAlaLeuGlyGlyPheGlu 415  
 Db 1281 GGAAGATCACTGAGGAGCAAGCTGAAGCCCATCTGGAATGGGCCCAAGGTGGCTTCAAG 1340  
 Oy 416 ProSerGlyProIysGlyLeuGluPro----- 424  
 Db 1341 CTTACTGGGATCGAGGCCCTCAAAACCACAGAAGCAACCAAGTGTAAATGTCGAAG 1400  
 Oy 425 -----ProGluGluGlyIysAsnProTyrIysGlu 434  
 Db 1401 GTGGGTGCTTCAGACAGTAAAGAACTTAAGAACCCAGAGACGGAGAAACAAAGTGAAGA 1460  
 Oy 435 ValTyrThrAspMetTyrValGluProGluAlaIleAlaIleTyrAlaProProProAla 454  
 Db 1461 GCACAAACCAAT-----GACTGTGCTGTTCATGATGCTCCCAACCAAG 1505

QY	445	LYL	YS	PRO	AG	YS	SER	TH	R	A	G	L	Y	S	PRO	LY	S	VAL	Y	S	G	L	I	L	A	S	G	L	Y	474				
DB	1506	C	G	C	T	C	A	A	G	A	C	A	A	T	A	C	T	A	T	G	C	G	G	A	A	G	A	G	G	A	G	1556		
QY	475	TH	R	AG	GU	AR	GU	LE	U	VAL	Y	R	G	U	VAL	A	R	G	U	L	Y	S	C	Y	S	A	R	G	U	LE	U	494		
DB	1557	A	G	C	G	A	A	C	G	A	T	G	C	T	T	C	T	G	A	A	G	C	A	A	C	A	A	G	G	A	A	C	1616	
QY	495	I	L	S	E	R	C	Y	S	G	L	S	E	R	U	A	N	VAL	T	H	R	L	E	U	G	L	U	N	I	S	P	R	U	514
DB	1617	T	T	G	C	T	G	T	G	A	A	G	A	A	G	A	C	C	T	G	T	G	C	T	T	C	C	A	C	C	C	C	1676	
QY	515	G	L	I	N	E	N	C	Y	S	A	N	C	Y	S	P	H	E	U	G	L	Y	S	A	L	A	Y	R	G	I	N	T	Y	534
DB	1677	C	A	G	A	T	G	C	G	G	A	T	G	C	T	T	C	T	T	C	T	T	C	A	A	T	G	A	A	G	A	G	1736	
QY	535	S	E	R	T	Y	R	C	Y	S	T	H	I	E	C	Y	S	G	L	Y	A	R	G	U	VAL	L	E	U	M	E	T	Y	554	
DB	1737	T	C	T	A	C	T	G	C	A	C	C	G	T	C	T	T	G	A	G	G	C	C	T	G	A	A	C	T	G	C	T	1796	
QY	555	C	Y	A	R	G	U	C	Y	S	VAL	G	L	Y	C	Y	S	VAL	A	S	P	H	E	U	VAL	G	I	P	R	O	G	I	574	
DB	1797	T	G	C	A	G	A	T	C	T	T	G	T	G	A	G	A	T	G	T	G	A	G	A	G	A	G	A	G	A	G	A	1856	
QY	575	A	L	A	I	L	E	S	E	R	C	Y	S	P	R	O	T	P	A	N	C	Y	T	Y	M	E	T	Y	M	E	T	Y	594	
DB	1857	G	C	A	A	G	C	T	G	C	A	A	C	C	T	G	A	C	T	C	T	A	T	G	T	G	C	C	T	C	C	A	1916	
QY	595	L	E	U	A	R	G	A	R	G	U	VAL	T	R	P	R	O	S	E	R	A	R	G	U	M	E	T	P	H	E	P	H	614	
DB	1917	C	T	C	G	A	G	C	A	A	G	A	A	T	T	G	A	A	C	A	T	T	T	C	T	T	C	A	A	C	T	T	1976	
QY	615	--	G	I	N	G	U	N	P	H	E	A	S	P	R	O	P	O	L	Y	S	VAL	T	Y	P	R	O	P	R	O	V	A	633	
DB	1977	C	T	G	A	A	G	A	T	T	G	A	C	C	A	G	T	T	G	A	C	C	A	G	A	T	T	C	T	G	A	G	2036	
QY	634	I	L	E	A	R	G	U	VAL	L	E	S	E	R	U	P	H	E	A	S	P	R	G	L	Y	L	E	U	VAL	L	E	U	653	
DB	2037	A	T	T	A	G	A	G	C	T	G	T	C	T	G	T	T	G	A	A	T	T	G	A	A	C	G	G	T	A	C	A	2096	
QY	654	G	L	I	L	E	G	U	VAL	A	S	P	R	G	U	VAL	C	Y	S	G	L	U	A	S	P	R	E	T	H	E	R	673		
DB	2097	G	G	A	T	T	A	A	G	A	A	G	T	A	C	A	T	T	G	C	C	G	A	A	G	T	T	G	C	A	G	A	2156	
QY	67																																	

```

/translation="MKGDSRHLNBEAGSGVBEICIIWNGNFSQSDSTKQAPPPVL
ALICEPTETRRGRSSKREVSLLNLTQDMTGDGRDDEVDGSDGLMPK
LURETKOTRTPSSPAVRTRHNSGTQSRAPRI.TKRGQRHHVGVDPVEAFPR
SRRRASSASSTWSPSPALGDMVKGI.KFGEWMPAMVYSKATSKQAMGPMRWQMDGK
GUSTEYQDKEFGIDGLVWGKI.KFGEWMPAMVYSKATSKQAMGPMRWQMDGK
SEI.SADMLVALGLFSGHFNLAT.FNKLVSYRKAMHYHLEKARVRAGKTFTSSPES
QLKPMLEWAHGGFKPTGTEGLUKPNKKQPVVVKSKVRSSRNLEPRRENKSRRTTN
DSAAEPSPPKRLKTNYSYGGKORDEDESESRMASEVTNNKGNLEDRCLSCGCKNPNV
PHAFEPGLGQCRDRFLPDELYMDYEDGQSYCTVCCEGRELLLCNNTSCCRFCVVC
LEVLVAGTADAEKAPLQSPWSCVMCLPQRCHGVLRRRKDNMRLQDFTTDDLEEFEP
PKLYPAIPAAKRPRIPRVLBEEMPFDLVIGSGPCNDLSNVNPAKGLVEGTGRFPF
EQGLKYNDVRKTIKGNIBEMGPFDLVIGSGPCNDLSNVNPAKGLVEGTGRFPF
HLLNYPTRKEDGNRPFPWFENVAMKVNDKDISRFACNPVIMDAIKVSAHAHRY
FWGNLPGKNRPMVMAKNDKLELQDCLFSTAKLKKVQITTTKNSIRQGNOLJFVW
MNGKDDVLWCTELERIFGPPAHYITDVSNMGRGARQKLLGRSSWSVFVIRHLFAPLKD
ACE"

BASE COUNT 1061 a 1059 c 1125 g 950 t
ORIGIN

Alignment Scores:
Pred. No.: 2,59e-80 Length: 4195
Score: 2221.00 Matches: 456
Percent Similarity: 58.32% Conservative: 126
Best Local Similarity: 45.69% Mismatches: 236
Query Match: 45.01% Indels: 160
Da: 10 Gaps: 15

US-09-720-086-7 (1-912) x AF068626 (1-4195)

QY 16 AlaAlaGluArgGluGluAspArgLyAspGlyGluGluGlnGluProArgGlyLyS 35
Db 23 GCGGCCCAAGTAAACGTAGCCAGCATCGCGCCGGAGATTCGCGAACCC-----GAC 76
QY 36 GluGluArgGlnGluProSerThrThrAlaArgLyAsValGlyArgProGlyArgLySArg 55
Db 77 ACTCCGGCGGCCCGCCCGCAGACGACCGCGCGCGATCGCGCGCCGCGCTACAGCCAG 136
QY 56 LysHisPro---ProValGluSerGlyAspThrProLyAspProAlaValIle----- 72
Db 137 CCTCAGCACAGGCCCGCGCTGAGGCTTTGGCCAGACCTTGGAACCTCAGGTATATACCTTT 196
QY 73 -----SerLySerProSerMet 78
Db 197 CGAGACGCGGGATCTCCCTCCCATCATGCTGTGGACCAANTCCAGGCGCTTC 256
QY 79 AlaGluAsp-----Sergly 83
Db 257 TTTCCGAACAAATGAGGGAGACAGCACACATCTCTGAATGAAGAGGGTGCCAGCGGG 316
QY 84 AlaSerGlu----LeuLeuProAsnGlyAspLeuGlyLyAsArgSerGluProGlnProGlu 102
Db 317 TATCAGAGGTGCATATGTTATGGGAACCTTCAGTGACCACTCTCAGACACGAAGGAT 376
QY 103 GluGlySerProAlaGlyGlyGlnLySgLyGlyAlaProAlaGluGlyGluGlyAlaAla 122
Db 377 GCTCCCTCACCC----- 388
QY 123 GluThrLeuProGluAlaSerArgAlaValGluAsnGlyCysCysThrProLyS----- 140
Db 389 -----CCAGTCTTGGGGAANTCTGCACAGACCGACGCTCTGCACACACAGACCA 439
QY 140 ----- 140
Db 440 GGCCGACAGGTCAAGCTCCCGCTCTCTTACAGAGGAGGTCTCCAGCCTTCTGAATTACACG 499
QY 141 -----GluGlyArgGlyAlaPro 146
Db 500 CAGGACATGACAGGAGATGGACAGACAGATGATGAAGTAGATGATGGGAATGCTCTGAT 559
QY 147 AlaGluAlaGlyLySgGluGlnLySgGluThrAsnIleGluSerMetLySgMetGluGlySer 166
Db 560 ATTCTAATGCCAAAGCTCACCCGTGAGACCAACAGACACCGAGCGCCTCTGGAAGCCCG 619

```

QY 167 ArgGlyArgLeuArgGlyGlyLeuGlyTyrProGluSerSerLeuArgGlyArgProMetPro 186  
 Db 620 GCTGTCCGAACCCGACATAGCAATGGAGCTCCAGCTTGAGAGCAAGAGCTCCGCC 679  
 QY 187 ArgLeuThr-----PheGlnAlaGlyAspPro 195  
 Db 680 AGAATCACCCGAGTCGGCAGGCGCCGACCATGTGCAGAGTACCTGTGGAGCTTTCCG 739  
 QY 196 TyrTyrIleSerLysArgLysArgAspGluTyrProLeuAlaArgTyrLysArgGluAlaGlu 215  
 Db 740 GCTACACAGGTCTCGAGACGTCGA----- 763  
 QY 216 LysLysAlaLysValIleAlaGlyMetAsnAlaValGluGluAsnGlnIleProGlyGlu 235  
 Db 764 -----GCATCGCTTCAGCAGCAAGCAGCCATGGTTCATCCCTGSCAGC 805  
 QY 236 SerGlnLysValGluGluAlaSerProProAlaValGlnGlnProThrAspProAlaSer 255  
 Db 806 GTGCACTTCATGAGAAAGTGAAGCTTAAGAGCGTCACTACCCCA-----TCA 853  
 QY 256 ProThrValAlaThrThrProGluProValGlySerAspAlaGlyAspLysAsnAlaThr 275  
 Db 854 GTTGACTTGACCGACAGATGAGATCAGAGGGTATGATACCAACAGGTGATGCAAG 913  
 QY 276 LysAlaGlyAspAspGluProGluTyrGluAspGlyArgGlyPheGlyIleGlyGluLeu 295  
 Db 914 AGCAGAGATGAGACAGACAGAGATATCAGGATGATAAGATTGGAATAGGTACCTC 973  
 QY 296 ValTyrGlyLysLeuArgGlyPheSerTyrProGlyArgGlyLevalSerTyrTyrMet 315  
 Db 974 GTGTGGGAAAGATCTCAGGGCTTCTCCGTGGCTGCATGCTGTGTCTCGAAGACC 1033  
 QY 316 ThrGlyArgSerArgAlaAlaGluGlyThrArgTyrValMetTyrPheGlyAspGlyLys 335  
 Db 1034 ACCTTCAGCGACAGGCCATGCGGAATGCGCTGGTACAGTGTGGTATGGCAAG 1093  
 QY 336 PheSerValValCysValGlyLysLeuMetProLeuSerSerPheCysSerAlaPheHis 355  
 Db 1094 TTTTCTGAGATCTGCTGACAACTGGTGGCTGTGGGGCTGTTCAGCGACACTTAAAT 1153  
 QY 356 GlnAlaThrTyrAsnLysGlnProMetTyrArgLysAlaIleTyrGluValLeuGlnVal 375  
 Db 1154 CTGGTACCTTCATATACCTGTTCTTATAGAAAGGCCATGTACACACTCTGAGAAA 1213  
 QY 376 AlaserSerArgAlaGlyLysLeuPheProValCysHisAspSerAspGluSerAspThr 395  
 Db 1214 GCCAGGTTCCAGGCTGGCAAGCCTTC-----TCCAGCACTCT 1252  
 QY 396 AlaLysAlaValGluValGlnAsnLysProMetIleGluTyrPalaLeuGlyGlyPheGln 415  
 Db 1253 GGAGAGTCACTGGAGGAGCAGCTGAAGCCCATGCTGGAGTGGCCACAGGTGGCTTCAAG 1312  
 QY 416 ProSerGlyProLysGlyLeuGluPro----- 424  
 Db 1313 CCTACTGGGATCGAGGGCTCAACCAAGCAAGCAACAGTGTAAATAGTCGAAG 1372  
 QY 425 -----ProGluGluGluLysAsnProTyrIleGlu 434  
 Db 1373 GTGCGTCTGTACAGACATGAAGAACTTAAGAACCGGAGACCGGAAACAAAGTGAAGA 1432  
 QY 435 ValTyrThrAspMetTyrValGluProGluAlaAlaIleTyrAlaProProProProAla 454  
 Db 1433 CGCAACAACCAAT-----GACTGTGCTGCTTCGATGCCCCCAACCCCAAG 1477  
 QY 455 LysLysProArgLysSerThrAlaGluLysProLysValLysGluIleIleAspGluArg 474  
 Db 1478 CGCCTCAAGACAAATAGCTATGCGGGAAGGACCGAGGGAG-----GATAGGAG 1528  
 QY 475 ThrArgGluArgLeuValTyrGluValArgGluLysCysArgAsnIleGluAspIleCys 494  
 Db 1529 AGCCGAAAGCGATGCTTCTTAAGTCAACCAACAAAGGCGATCTGGAAACGCGCTGT 1588  
 QY 495 IleSerCysGlySerLeuAsnValThrLeuGluHisProLeuPheValGlyGlyMetCys 514

Db 1589 TTGTCTGTGGAAAGAAACCCCTGTGTCTTCACACCCCTTTTGAGGGGTCTGT 1648  
 QY 515 GlnAsnCysLysAsnCysPheLeuGluCysAlaIleTyrGlnTyrAspAspAspGlyTyrGln 534  
 Db 1649 CAGAGTTCGGGATCCCTTCTTAAGACTCTTTTACATGTAATGATGAGACGGCTTAACAG 1708  
 QY 535 SerTyrCysThrIleCysCysGlyGlyArgGluValLeuMetCysGlyLysAsnAsnAsnCys 554  
 Db 1709 TCTTACTGACACCGTGTCTGTGAGGGCGGTGAACGTGCTGTGCATTAACACAGCTGC 1768  
 QY 555 CysArgCysPheCysValGluCysValAspLeuLeuValGlyProGlyAlaIleAlaGlnAla 574  
 Db 1769 TGCAGATGCTTCTGTGTGAGTGTCTGAGAGTGTGTGGGGCGGACGACAGCTGAGAGAT 1828  
 QY 575 AlaIleLysGluAspProTyrAsnCysTyrMetCysGlyLysLysGlyTyrTyrGlyLeu 594  
 Db 1829 GCCAAGCTGCAGAAACCTGTGATGCTATATGTGCTCCCTCAGCCCTGCATGGGGGTG 1888  
 QY 595 LeuArgArgArgGluAspTyrProSerArgLeuGlnMetPhePheAlaAsnAsnHisAsp 614  
 Db 1889 CTCGAGCGCAGGAAGATTGGACATGCGCTGCAAGACTTCTTCACTAGTATCCTGAC 1948  
 QY 615 ---GlnGluPheAspProProLysValTyrProProValProAlaGluLysArgLysPro 633  
 Db 1949 CTGGAAAGATTGAGCCACCAAGTGTATACCCCAATTCCTGCAGCCMAAGAGAGGCC 2008  
 QY 634 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeuLysAspLeu 653  
 Db 2009 ATTGAGCTCTGTCTGTGTGTAAGAAATGCAAGGGGTACTTGGTGTCAAGAGTTG 2068  
 QY 654 GlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIleThrValGly 673  
 Db 2069 GGTATTAAAGTGAAGAAAGTACATTGCTCCGAACTCTGTGAGAGTCAATCGCTGTGGGA 2128  
 QY 674 MetValArgHisGlnGlyLysIleMetTyrValGlyAspValArgSerValThrGlnLys 693  
 Db 2129 ACTGTAAAGCATGAAGGCCAGATCAATATGTCAATGACGTCGGAATATCACAAGAAA 2188  
 QY 694 HistLeuGlnIleTyrProPheAspLeuValIleGlyGlySerProCysAsnAspLeu 713  
 Db 2189 AATATTAAGAGTGGGGCCGCTGTGACTGTGTGATGGTGAACCCCATGGAATGATCTC 2248  
 QY 714 SerIleValAspProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhePheGlu 733  
 Db 2249 TCTTAAGTCAATCTGCGCCGGAAGGTTTATATAGGGCACAGAAAGCTTTCTTCAG 2308  
 QY 734 PheTyrArgLeuLeuHisAspAlaArgProLysGluGlyAspAspArgProPhePheTyr 753  
 Db 2309 TTTTACCACTGTGAAATTATATACCGCCCAAGAGGGCGACAACGTCATCTTCTGG 2368  
 QY 754 LeuPheGluAsnValValAlaMetGlyValSerAspLysAspAspIleSerArgPheLeu 773  
 Db 2369 ATGTTCGAGATGTGTGGCCATGAAGTGAATGACAAAGAAAGCATCTCAAGATTCTGT 2428  
 QY 774 GluSerAspProAlaMetIleAspAlaLysGluValSerAlaIleHisArgAlaArgTyr 793  
 Db 2429 GCATGTAAACCAAGTATGATGATGCCATCAAGGTGTGTGTCTACAGGGCCCGGTAC 2488  
 QY 794 PheTyrGlyAsnLeuProGlyMetAsnArgProLeuAlaSerThrValAsnAspLysLeu 813  
 Db 2489 TTTCTGGGTAACTTACCGGAAATGAACAGGCCGTGATGCTTCAAGAAATGATAGCTC 2548  
 QY 814 GluLeuGlnGluCysLeuGluHisGlyArgIleAlaLysPheSerLysValArgThrIle 833  
 Db 2549 GAGCTGCAGACTCTGTGAGTTCAGTGAAGCAGCAAAAGTTAAAGAAAGTCAACACATA 2608  
 QY 834 ThrThrArgSerSerAsnSerIleLysGlnGlyLysAspGlnHisPheProValPheMetAsn 853  
 Db 2609 ACCACCAAGTCAACTCCATAGACAGGGGCAAAACAGCTTTTCCCTGTATGTATGAAT 2668  
 QY 854 GluLysGluAspIleLeuTyrCysThrGluMetGluArgValPheGlyPheProValHis 873

```
Db 2669 GCAAGGACGACGCTTTTGTGTGCACTGAGCTCGAAGAGTCTTCGGCTTCCCTGCTCAC 2728
Qy 874 TyrThsPValSerAsnMetSerArgLeuAlaArgGlnArgLeuLeuGlyArgSerTrp 893
Db 2729 TACACGAGCTGTCAACATGGCGCGCGCGCTCAGAAAGCTGCTGGCAGGCTCTGG 2788
Qy 894 SerValProValIleArgHisLeuPheAlaProLeuLysGlyLysTrpPheAlaCys 911
Db 2789 AGTGATACCGGTATCAGACACCTGTTTGGCCCTTGAAGGACTACTTTCCTGT 2842

RESULT 13
AF151969
LOCUS AF151969 4338 bp mRNA linear ROD 08-JUN-2000
DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 1
(Mnmt3b) mRNA, complete cds.
ACCESSION AF151969
VERSION AF151969.1 GI:8347117
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4338)
AUTHORS Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
Shen,Y.
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing
isoforms in mouse embryonic tissue
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4338)
AUTHORS Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
Shen,Y.
DIRECT SUBMISSION
TITLE Submitted (17-MAY-1999) Department of Biochemistry and Molecular
Biology, Institution of Basic Medical Sciences, Chinese Academy of
Medical Sciences, Dong Dan San Tiao 5, Beijing city 100005,
P.R.China
FEATURES
source
1..4338
/organism="Mus musculus"
/strain="KM"
/db_xref="taxon:10090"
/dev_stage="8-9 day old embryo"
1..4338
/gene="Dnmt3b"
412..2991
/gene="Dnmt3b"
/functions="methylates cytosine in DNA"
/notes="alternatively spliced"
/codon_start=1
/product="DNA cytosine-specific methyltransferase isoform
1"
/protein_id="AAF74515.1"
/db_xref="GI:8347118"
/tranlation="MKGDSRHLEEGAGGYEECIIVNGNFDSDSDTKDAPSPVPLE
AICTEPVCTPETSRESSRLSKREVSLNLYTODMTGDDRDDEVDGSDILMPK
LTRETQDTRTESPAVTRHSNCTSLERQASPRITRGROGRHHVQEPVUEPNTF
SRBRASSASTWSSPASVDMEVTHKSVSTPSVDLSODGQEGMDTIQVDAESIY
GDSTEQDKEFGIGLWVKLGFSGWMPAVVSWKATSKRQMPGRWVQWGGAF
SEISADKLVALGUFSGHFNLATFNKLVSIRKAMHTLEKARVAGTKITSSGSELD
QKPMLEWARGPKPTGIEGLAPNKQPVNKSVDSDNLEPRRNSKRRRTIN
DSKASEPPPKRLKTNYSYGDGDEEERSEWSEVNNKNGLEDRLCSCKKNPV
PFLPVEGGLQCSDRDFELFYMYDEDDGYOYCTVCCGREGLLCSNTSCRCFCFEC
LEVLVAGTAEDAKQEPWSCYMLPQRCGVLRRRKDNMRLLQDFFTDPLEEFEP
KGLYPALPAAKRPIRVLSLFDGIATGLVLELGIKVEKYIASEVCAESIATGVFK
EQGIKYVNDVRKTKNIIEWGPFDLVIGSPCNDLSNVNPAKRGLEYGTGRFFEFY
HLNLTYPKPEGDNRPFFMNFVNMVKNVKKDTSRLACNPVMDAIKVSAAHFRAY
FWGNLPMNRPNVMSKNDKLEQCLFEFRTAKLKKVQITITTKNSIRQGNKLPVY
MNGKDDVLWCTELERIFGPAPHYITDVSNMGRGARQKLLGRSWSVPVIRHLPKDYF
ACE"
BASE COUNT 1073 a 1124 c 1173 g 968 t
ORIGIN
Alignment Scores:
```

```
Pred: No.: 2,93e-80 Length: 4338
Score: 2220.00 Matches: 462
Percent Similarity: 57.87% Conservative: 126
Best Local Similarity: 45.47% Mismatches: 253
Query Match: 44.98% Indels: 175
DB: 10 Gaps: 17

US-09-720-086-7 (11-912) x AF151969 (1-4338)

Qy 2 ProAlaMetProSer--SerGlyProGlyAspThrSerSerSer-----AlaAlaGlu 18
Db 147 CCGGACGCGCGGGTTAAGCCGCCCAAGTAAACGTAGCCGAGATCGCGCGCGGAGAT 206
Qy 19 -ArgGluGlnAspArgGlyAspGlyGluGlnGluGluProArgGlyLysGluGlnAr 38
Db 207 TCGCGAACCCGAC-----ACTCCGCG 227
Qy 38 gGlnGluProSerThrThraAlaArgLysValGlyValArgProGlyArgLysHisPr 58
Db 228 CCGCCCGCGCGCGAGGACCGCGCGCGATCGCGCGCGCTTACAGCCAGCCTCACGA 287
Qy 58 o----ProValGluSerGlyAspThrProLysAspProAlaValIle----- 72
Db 288 CAGGCGCGCTGAGGCTTGTGCCAGACCTTGGAAACCTCAGGTATATACCTTTCCAGC 347
Qy 73 -----SerLysSerProSerMetAlaGlnAs 81
Db 348 GGGATCTCCCTCCCTCCATCATGTGCTTGGGACCAATCCAGGCGCTTCTTTCA 407
Qy 81 p-----SerGlyAlaSerGl 86
Db 408 AACATGAGGGAGACAGCAGCATCTGANTAGAGAGAGGGTCCAGCGGTATGAGGA 467
Qy 86 u----LeuLeuProAsnGlyAspLeuGlnLysArgSerGluProGlnProGluGlySe 105
Db 468 GTGCATATCTTAAATGGAGACTTCAGTGACCACTTCAGACCAAGAGATGCTCCCTC 527
Qy 105 rProAlaGlyGlyGlnLysGlyGlyAlaProAlaGluGlyGluGlyAlaAlaGluThr 125
Db 528 ACCC----- 531
Qy 125 uProGluAlaSerArgAlaValGluAsnGlyCysCysThrProLys----- 140
Db 532 -CCAGTCTTGGAGGCAATCTGCACAGAGCCAGTCTGCACACAGACAGACGCGCAG 590
Qy 140 ----- 140
Db 591 GTCAAGCTCCCGGCTGTCTAAGAGGGAGGTCTCCAGGCTTCTGAATTACACGAG 650
Qy 141 -----GluGlyArgGlyAlaProAlaGluAl 149
Db 651 GACAGGAGATGGAGACAGAGATGATGAAGTAGATGGGAATGGCTCTGATATTCTA 710
Qy 149 aGlyGlyGluGlnLysGluThrAsnIleGluSerMetLysMetGluGlySerArgGlyAr 169
Db 711 GCCAAAGCTACCGTGAGACCAAGGACACAGGACCGCGCTCTGAAAGCCCGGCTG 770
Qy 169 gLeuArgGlyGlyLeuGlyTrpGluSerSerLeuArgGlnArgProMetProArgLeuTh 189
Db 771 AACCCGACATAGCAATGGGACCTCCAGCTTGGAGAGCAAGAGACCTCCCCAGAA 830
Qy 189 r-----PheGlnAlaGlyAspProTyTrI 198
Db 831 CCGAGGTCCGCGCGCGCCACCATGTGCGAGGAGTACCTCTGTGGAGTTTCCGGCTAC 890
Qy 198 eSerLysArgLysArgAspGluTrpLeuAlaArgTrpLysArgGluAlaGlyLysAl 218
Db 891 GTCTCGGAGACGTGCA----- 906
Qy 218 aLysValIleAlaGlyMetAsnAlaValGluGluAsnGlnGlyProGlyGluSerGln 238
Db 907 -----GCATCTCTTCCAGCAAGCAGCAGCCCATGTCTATCCCTGCCAGCGTCGACT 956
```

QY 238 eValGluGluAlaSerProProAlaValGlnGlnProThrasProAlaSerProThra 258  
 Db 957 CATGAAAGAAATGACCTTAAGACCTGACATACCA-----TCAGTTGACTT 1004  
 QY 258 lAlaThrThrProGluProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaG1 278  
 Db 1005 GAGCCAGAGATGAGATACAGAGGGTATGGATACACACAGGTGATGACAGACATATA 1064  
 QY 278 yAspAspGluProGluLysArgGlyLysArgGlyPheGlyLysGlyLysLeuValTrpG1 298  
 Db 1065 TGGAGACAGACAGATATACAGATGATTAAGAGTTTGGAATAGTAGCCTCGTGTGGG 1124  
 QY 298 yLysLeuArgGlyPheSerTrpTrpProGlyLysAlaLysSerTrpTrpMetThrGlyAr 318  
 Db 1125 AAATATAGAGGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1184  
 QY 318 gSerArgAlaAlaGluGlyThrArgTrpValMetTrpPheGlyLysArgGlyLysPheSerVa 338  
 Db 1185 GCGACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1244  
 QY 338 lValCysValGluLysLeuMetProLysSerSerPheCysSerAlaPheHISGlnAlaTh 358  
 Db 1245 GATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1304  
 QY 358 rTyAsnLysGlnProMetTyArgLysAlaLysArgGlyLysLeuGlnValAlaSerSe 378  
 Db 1305 CTTCATATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1364  
 QY 378 rArgAlaGlyLysLeuPheProValCysHisAspSerAspLysSerPheTrpAlaLysAl 398  
 Db 1365 TCGAGCTGGCAAGCTTC-----TCAGCAGCTCTGAGAGCTC 1403  
 QY 398 aValGlnValGlnAsnLysProMetLysGluTrpAlaLeuGlyLysPheGlnProSerG1 418  
 Db 1404 ACTGAGAGACCAAGCTGGAAGCCATGCTGAGATGGGCCCAAGCTGCTCAAGCTTACTG 1463  
 QY 418 yProLysGlyLeuGluPro-----424  
 Db 1464 GATGAGGCTCTCAAAACCAACAAAGCAACAGCTGCTAATAAGTGAAGTGGCTGCTG 1523  
 QY 425 -----ProGluGluGluLysAsnProTyLysGluValTyTh 437  
 Db 1524 TTCAGACAGTAGAAGCTTGAAGCCAGAGAGCGGAGAAACAAAGTCCAGAACGACACAC 1583  
 QY 437 rAspMetTrpValGluProGluAlaAlaAlaTyArgAlaProProProAlaLysLysPr 457  
 Db 1584 CAAT-----GACTCTGCTGCTTGCAGTCCCTCCCAAGCCCTTAA 1628  
 QY 457 cArgLysSerThrAlaGluLysProLysValLysGluLysLeuAspGluArgThrArgG1 477  
 Db 1629 GACAAATATAGCTATGGCGGAGAGACCGAGGGAG-----GATGAGAGAGCGGAG 1679  
 QY 477 uArgLeuValTyArgLysAlaArgGlnLysCysArgAsnLysGluAspLysCysLysSerCy 497  
 Db 1680 ACCGATGGCTTCTGAAGTCAACCAACAAAGGCAATCTGGAAGACCTCTTTGCTCTG 1739  
 QY 497 sGluSerLeuAsnValThrLeuGluHisProLeuPheValGlyGlyMetCysGlnAsnCy 517  
 Db 1740 TGGAAAGAGAAACCTGCTGCTTCCACCCCTCTTAAGAGTGGGCTCTGCAAGAGTTG 1799  
 QY 517 sLysAsnCyPheLeuGluCysAlaTyArgLysTyArgAspAspGlyTyArgLysSerTyCy 537  
 Db 1800 CCGGATCGCTTCTTAAGAGCTCTTCAATATAGATAGAGACCGCTATACAGTCTTACTG 1859  
 QY 537 sThrLysCysGlyGlyArgGluValLeuMetCysGlyLysAsnAsnAsnCyCysArgCy 557  
 Db 1860 CACGCTGCTGCTGAGGGCCCTGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1919  
 QY 557 sPheCysValGluCysValAspLeuLeuValGlyProGluAlaAlaAlaAlaLysLys 577  
 Db 1920 CTTCCTGCTGAGAGTCTGAGAGTCTGAGTGGGCGCAGAGCTGAGTGGAGTGGCAAGCT 1979  
 QY 577 sGluAspProTrpAsnCyTyMetCysGlyHisLysGlyThrTyArgLysLeuLeuArgAr 597

Db 1980 GCAGAAACCTCGAGCTGCTATATGCTCTCCCTCACGGCTGCCATGGGGCTCTCCACG 2039  
 QY 597 gArgGluAspTrpProSerArgLeuGlnMetPhePheAlaAsnAsnHisAsp---GlnG1 616  
 Db 2040 CAGAAAGATTTGAAACATGCGCTCGCAAGACTTCTTCACTACTGATCTCTGACCTGGAAGA 2099  
 QY 616 uPheAspProProLysValTyArgProValProAlaGluLysArgLysProLysArgVa 636  
 Db 2100 ATTGAGCCACCCCAAGTTGATCCAGCAATCTCGACAGCCCAAAAGAGGCGCATTAAGT 2159  
 QY 636 lLeuSerLeuPheAspGlyLysAlaThrGlyLeuLeuValLeuLysAspLeuGlyLysG1 656  
 Db 2160 CCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2219  
 QY 656 nValAspArgTyTrpLeuLysSerGluValCysGluAspSerLysPheValGlyMetValAr 676  
 Db 2220 AGTGAAGAGTACATTCCTCTCCAGAGTCTGTGACAGTCCATCCCTGTGGAACTTTAA 2279  
 QY 676 gHisGlnGlyLysLysMetTyArgLysAspValArgSerValThrGlnLysHisLysG1 696  
 Db 2280 GCATGAAGGCGCAGATCAATATGTCATGACGTCCGGAATAATCAACAAAGAAATATTGA 2339  
 QY 696 nGluTyArgLysProPheAspLeuValLysGlyLysProCysAsnAspLeuSerLysVa 716  
 Db 2340 AGAGTGGGGCCGCTTGCATCTGATGATGGTGAAGCCCATGCAATGATCTCTTAAGT 2399  
 QY 716 lAsnProAlaArgLysGlyLeuTyArgLysGlyTyArgLysLeuPheGlyLysPheTyAr 736  
 Db 2400 CAATCTGCTGCGCAAGAGTTATATAGGCGCACAGAAAGCTCTTCTCGAGTTTAAACA 2459  
 QY 736 gLeuLeuHisAspAlaArgProLysGluGlyAspAspArgProPhePheTrpLeuPheG1 756  
 Db 2460 CTTCCTGATATATACCGGCCCAAGAGAGGCGGACACCGTCCATCTTCTGAGTCTTGA 2519  
 QY 756 uAsnValValAlaMetGlyValSerAspLysArgAspLysSerArgPheLeuGluSerAs 776  
 Db 2520 GAATGTTGGCCATGAAGAGTAATGACAAAGAAAGATCATCTCAAGATTCTCGCATGAA 2579  
 QY 776 nProValMetLysAspAlaLysGluValSerAlaAlaHisArgLysAlaArgTyPheTrpG1 796  
 Db 2580 CCCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2639  
 QY 796 yAsnLeuProGlyMetAsnArgProLeuAlaSerThrValAsnAspLysLeuGluLeuG1 816  
 Db 2640 TAACTTACCGGAAATGAACAGCCCGTGAAGCTTCAAAAGATATATAGCTCGAGCTGCA 2699  
 QY 816 nGluCysLeuGluHisGlyArgLysAlaLysPheSerLysValArgThrLysThrArg 836  
 Db 2700 GACTGCTGAGTTCAAGTATGAGACAGCAAGAGTTAAAGAAAGTGCAGCAATTAACCAACA 2759  
 QY 836 gSerAsnSerLysLeuGlnGlyLysAspGlnHisPheProValAlaPheMetLysGluLysG1 856  
 Db 2760 GTCAAGCTTCAATCAAGAGGCAAAACCAAGCTTCTCTGTATCTATGATGCAAGGA 2819  
 QY 856 uAspLysLeuTrpCysThrGluMetGluArgValPheGlyPheProValHisTyThrAs 876  
 Db 2820 CGAGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2879  
 QY 876 pValSerAsnMetSerArgLeuAlaArgGlnArgLeuLeuGlyLysSerTrpSerValPr 896  
 Db 2880 CGTGTCCAACTGGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2939  
 QY 896 oValLysArgHisLeuPheAlaProLeuLysGluTyPheAlaLys 911  
 Db 2940 GGTATCAGACCACTGTTGGCCCTTGAAGAGTACTTGGCTGT 2985  
 RESULT 14  
 AF331857 AF331857 4335 bp mRNA linear PRI 02-JAN-2002  
 LOCUS Homo sapiens DNA cytosine methyltransferase 3 beta (DNMT3B) mRNA,  
 complete cds.  
 ACCESSION AF331857

```

VERSION AF331857.1 GI:18033254
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 4335)
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          2 (bases 1 to 4335)
          Ni, J., Pradhan, S. and Roberts, R.J.
          Cloning, expression and characterization of human DNMT3 genes
          Unpublished
REFERENCE 1 (bases 1 to 4335)
          Ni, J., Pradhan, S. and Roberts, R.J.
          Direct Submission
          Submitted (22-DEC-2000) New England Biolabs, 32 Tozer Road,
          Beverly, MA 01915, USA
FEATURES
     source
     1..4335
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
     gene
     1..4335
        /gene="DNMT3B"
     CDS
     322..2883
        /gene="DNMT3B"
        /codon_start=1
        /product="DNA cytosine methyltransferase 3 beta"
        /protein_id="AAL57040.1"
        /db_xref="GI:18033255"
        /translation="MKGDRHLNGEEDAGGREDGILVNGACSDQSDSPPILEARTP
        EIRGRSSRLSKREVSSLLSYTDLTGDDGDEGSDTPVMPKLPFRPTRISSPA
        VTRNNSSSRERHPRSPRSTGQGNHVDSPVFPATRSIPRERATASAGTPWS
        PPSYLITDITDTEHTGTPSSSTPYARLAODSOOGMESPOVEDGDDGSEVO
        DKRFGIGDLVWGKIKGFSWPMVVMVKATSKPOAMGNBVOMFGDGKSEVADK
        LVALGLESQHENLATEPKLYSKMYHALKARVAGTTPSSRGDSLEDLKPMLE
        WAHGFFPTGIEGLFNNTQPVYKSKVRAGSRKLESRYENKTRRRTRDDSATSDY
        CFPAPKUTKCNNGDQDEQREQNASDVANNKSLSDGCLSCRNKPNVSPFPLF
        EGGCLTDRFLFLEFYIDDDGYOICTVCCGRELILCSNCCRCFCVCELEVLV
        GTGAARAKLQEPWSCYMCIPQRCVGLRRKQNMVNLQAFPTSDTGLYEAPKGIYA
        IPAAARRPIKLVSLFDGIATGLVLKELGKIVKIVYAVESEESIAVGRHEGNKRY
        VNDVNRITKKNIEWGFPLVIGVSPCNDLNNVPARKGLYEGTGRLEFFVHLNYS
        RPKEGDDRPKPFVFNFMVAMKVGDKRDISRFLECNPMVIDAIKVSAAHRAFYFNGNP
        GMPNPVJASKNDKLEQDCLEYNRIAKLKYQTITTKNSIKQGNQLFPVVMNGKD
        VLWCTELERIFGFPVHTVDSNMGRGARQKLLGRSNVSPVIRHLFAPLKDYPAE"
BASE COUNT 1045 a 1172 c 1167 g 951 t
ORIGIN
Alignment Scores:
Pred. No.: 6.69e-80 Length: 4335
Score: 2211.00 Matches: 477
Percent Similarity: 59.54% Conservative: 122
Best Local Similarity: 47.42% Mismatches: 263
Query Match: 44.80% Indels: 147
DB: 9 Gaps: 23

US-09-720-086-7 (1-912) x AF331857 (1-4335)
QY 2 ProAlaMetProSerSerGlyProGlyAspThrSerSerSerAlaAlaGluArgGluGlu 21
Db 8 CCGCTGCCCCG---TCCGGCCCGCCGCGCTTCTCCGACGAGCTGCTCC-CGGCTCCGC 63
QY 22 AspArgLysAspGlyGluGluGluGluGluGluGluGluGluGluGluGluGlu 40
Db 64 GCGCGC-----AGCCCGCGTGGACGCTCCGAGCGCCCGCCCGACG 102
QY 41 ProSerThrThrAlaArgLysValGlyArgProGlyArgLysArgLysHisProProVal 60
Db 103 GACGGGACCGGCTCTCCGCGGTGCGGCGAGCGGGCGCAACGCTGCGCGCGCGCGAGCG 162
QY 61 GluSerGlyAspThrProLysAspPro---AlaValIleSerLysSerProSerMetAla 79
Db 163 CTGGGGTTAAGTGGCCCAAGTAAACCTAGCTCGCGCATCGGCGCGGAGATTCCGAGCC 222
QY 80 Gln-----AppSerGlyAlaSerGluLeuLeuProAsnGlyAspLeuGlu 94

```

```

Db 223 CAGCGCCCTGCACGGCGCGCCAGCGCGCTTCCCGCCAGCCAGCCCGCCCGCGCTCCGC 282
QY 95 LysArgSerGluProGlnPro-----
Db 283 CCGCCAGCGCGCGCCCGCCAGCCAGCCCGCTTGGCGGAGGAAAGCATGAAGGAGGACACCGCAT 342
QY 102 -----GluGluGlySerProAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 113
Db 343 CTCATGTGAGAGGAG-----GACGCGCGCGGAGGAGGAGACTCGATCTCTCGTCAACGGG 396
QY 114 AlaProAlaGluGlyGluGlyAlaAlaGluThrLeuProGluAlaSerArgAlaValGlu 133
Db 397 GCGTCGACGACCGACGCTCTCCGACTCGCCCAATCTCTGGAGGCTATCCGC-----447
QY 134 AsnGlyCysCysThrProLys---GluGlyArgGlyAlaProAlaGluAlaGlyLysGlu 152
Db 448 -----ACCCCGGAGATCAGAGCGCCGAAGATCAAGCTGCGGACTCTCCAAGAGG 495
QY 153 Gln-----
Db 496 GAGGTGTCTCAGTCTCTAAAGCTACACAGGACTTTGACAGCGCATGCGCGGGAAGAT 555
QY 154 -----
Db 556 GGGGATGGCTCTGACACCCCGACGTCATGCGCAAGCTCTTCCGGGAAACAGGACTCGTTCA 615
QY 161 MetLysMetGluGlySerArgGlyArgLeuArgGlyGlyLeuGlyTrpGluSerSerLeu 180
Db 616 -----GAAAGCCAGCTGTCCGAACTCGAAATAACAACAGTGTCTCCAGCGGGAG 666
QY 181 ArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyrTyrIleSerLys 200
Db 667 AGGCACAGGCTTCCCACTTCCACCC--GAGGCCGCG-----703
QY 201 ArgLysArgAspGluTrpLeuAla---ArgTrp-----210
Db 704 AGGCGCGCAACCATGTGGAGAGTCCCCCGTGGAGTTCGCCGGCTACAGGTCCTCGAGAC 763
QY 211 -----LysArgGluAlaGluLysLysLysValIleAlaGlyMetAsnAlaVal 227
Db 764 GCGCGGCAACAGCATCGGAGGAGAGCCATGCGCGCTCCCTCCCGCTTCTTACCTACCA 823
QY 228 GluGluAsnGlnGlyProGlyGluSerGlnLysValGluGluAlaSerProProAlaVal 247
Db 824 TCGACCTTCACAG-----ACGACACAGAGACACATGGAGCGCCCGACGACGCA 874
QY 248 GlnGlnProThrAspProAlaSerProThrValAlaThr-----ThrProGlu 263
Db 875 GTACCCCTTAGC--CCCGCTAGCCAGGACGAGCGGCGGCGCATGGAGTCCCCGC 931
QY 264 ProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAspGluProGlu 283
Db 932 AG-GTGGNGGCACACAGTGGAGAT-----GGAGACAGTTCAGAG 969
QY 284 TyrGluAspGlyArgGlyPheGlyIleGlyLeuValTrpGlyLysLeuArgGlyPhe 303
Db 970 TATCAGGATGGGAAGGAGTTTGAATAGGAGCTCTGTTGGGGAAGATCAAGGCGCTTC 1029
QY 304 SerTrpTrpProGlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGlu 323
Db 1030 TCCTGTGGCGCGCCATGTGGTGTGTTTGGAGGCCACCTCCCAAGCGACAGGCTATGTCT 1089
QY 324 GlyThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLys 343
Db 1090 GGCATGCGGTGGGTCCAGTGGTGGTGGATGGCAAGTCTCCGAGGTCTCTGACAGACAA 1149
QY 344 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnPro 363
Db 1150 CTGTGGCACTTGGGCTGTTTACGACGAGCAGCTTTAATTTGGCCACCTTCAATAGCTCGTC 1209
QY 364 MetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerArgLysArgLysLeu 383
Db 1210 TCTATCGAAAGCCATGTACCATGCTCTGGAGGAAGCTAGGTTGGAGCTGCGAGACC 1269

```



QY 384 PheProValCysHisAspSerAspGluSerAspThrAlaValGluValGlnAsn 403  
 Db 1270 TTCCCC-----AGCAGCCCTGGAGACTCATTTGAGAGCACAGCTG 1308  
 QY 404 LysProMetIleGluTrpAlaLeuGlyIlePheGlnProSerGlyProIleGlyLeuGlu 423  
 Db 1309 AAGCCCATGTTGGAGTGGGCCCGACCGGGGCTTCAAGCCCACTGGATCGAGGGCTCTCAA 1368  
 QY 424 ProProGluGlu----- 427  
 Db 1369 CCCAACACACGACGACGACGCTGTTAATAGTCGAAGGTCCTGCTGACGACGTAGAAA 1428  
 QY 428 ---GluysAsnProTyrLysGluValTyrThrAspMetTrpValGluProGluAla 446  
 Db 1429 TTAGAAATCAAGAAATACAGAACAAAGACTCGAAGACGACAGCTGACAGACTGACCCACC 1488  
 QY 47 Ala---TyrAlaProProProProAlaIleLysProAlaGlySerThrAlaGluLysPro 465  
 Db 1489 TCTGACTACTGCTCCGACCCGACGCGCTCAAGACMAATTGCTATACACGCGCAAAAGAC 1548  
 QY 466 LysValLysGluIleIleAspGluThrArgGluArgGluValTyrGluValArgGln 485  
 Db 1549 CGAGGGATGA-----GATCAGACCGAGAACAAATGGCTTCAGATGTTCCAAC 1599  
 QY 486 LysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThrLeuGlu 505  
 Db 1600 AACAAAGACACCTGAGATAGCTGTTGTTGCTTGTGACAGAAACCCCGTGTCTTC 1659  
 QY 506 HisProLeuPheValGlyIleCysGlnIleAsnCysLysAsnCysPheLeuGluCysAla 525  
 Db 1660 CACCTCTCTTGAAGGGGGGCTGTGTGACATGCCCGGATCGCTTCTTGAAGCTGTTT 1719  
 QY 526 TyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyGlyValArgGlu 545  
 Db 1720 TACATGTAATGATGACGATGGCTATAGTCTTACATGCTGATGCTGGAGGCGGAGAG 1779  
 QY 546 ValLeuMetCysGlyAsnAsnAsnAsnCysCysArgCysPheCysValGluCysValAspLeu 565  
 Db 1780 CTGCTGCTTTCAGACACACGAGCTGCTGCGGCTTCTGTGTGATGCTGTGAGAGTG 1839  
 QY 566 LeuValGlyProGlyAlaIleGlnAlaIleLysGluAspProTyrAsnCysTyrMet 585  
 Db 1840 CTGGTGGGACACGAGCGCGCGCCGAGCTTCAAGCTTCAGAGCCCTGAGAGCTTACATG 1899  
 QY 586 CysGlyHisLysGlyTyrThrGlyLeuLeuArgArgGluAspTrpProSerArgLeu 605  
 Db 1900 TGTCTCCCGCAGCGCTGTCAATGCGCTCTGCGCGCGCGAGAGACTGGAACGTGCGCTG 1959  
 QY 606 GlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyrProPro 625  
 Db 1960 CAGGCTCTTTCACACGAGACACGCGGCTTGAATATGAAGCCCAACACTGTACCTGCC 2019  
 QY 626 ValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIleAlaThr 645  
 Db 2020 ATTCCCGCAGCGCCGAGCGCGCCCATTCGAGCTCTGATGTTGATGATGATCGCAGCA 2079  
 QY 646 GlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSerGluVal 665  
 Db 2080 GGGTACCTAGTCTCTCAAGAGATGGGCTAAAGTAGGAAGATAGTGTGCTTCTGAATG 2139  
 QY 666 CysGluAspSerIleThrValGlyMetValArgHisGlnGlyIleGlyMetTyrValGly 685  
 Db 2140 TGTGAGAGAGTCCATTCGCTTGGAAACCTGAAGCAGAGGGGAAATATCAATACCTGAAC 2199  
 QY 686 AspValAlaSerValThrGlnLysHisIleGlnGluTyrGlyProPheAspLeuValIle 705  
 Db 2200 GACGTGAGAGACATCAACAAGAAATATTTGAAGATGGGCGCCATTGACTTGGTGATT 2259  
 QY 706 GlyLysSerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeuTyrGlu 725  
 Db 2260 GCGGAGAGCCCATGCAAGATCTCAAAATGTAATCCAGCAGAAAGGCTGTATAG 2319

QY 726 GlyThrGlyValGluLeuPheGluPheTyrArgLeuLeuHisAspAlaArgProLysGlu 745  
 Db 2320 GGTACAGGCCCGCTCTTCTTCTGAAATTTTACACCTGTGAAATTTACTCAGCCCAAGAG 2379  
 QY 746 GlyAspAspArgProPheProPheTrpLeuPheGluAsnValAlaIleMetGlyValSerAsp 765  
 Db 2380 GGTATGACCGGCGCTTCTTCTTGTGATGTTGAGATGTTGAGCATGATGACATGAGCTGGCAC 2439  
 QY 766 LysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLysGluVal 785  
 Db 2440 AAGAGGACATCTACCGCTTCTCTGAGCTGATTCAGATGATGATGATGATGATGATGATGAT 2499  
 QY 786 SerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArgProLeu 805  
 Db 2500 TCTGCTGCTCACAGGGCCCGATCTTCTGAGGCAACTTACCAGGATGAAACAGGCCGCTG 2559  
 QY 806 AlaSerThrValAsnAspLysLeuGluLeuGlnIleCysLeuGluHisGlyValArgIleAla 825  
 Db 2560 ATACATCTAAAGAAATGATTAACCTGAGCTGACAGACTGCTTGAATATCAATAGATAGCC 2619  
 QY 826 LysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGlyLysAsp 845  
 Db 2620 AAGTTAAAGAAATGACACACATTAACCAAGTCGAATCTGATCAACACAGGAAAC 2679  
 QY 846 GlnHisPheProValPheMetAsnGluLysGluAspIleLeuTyrCysThrGluMetGlu 865  
 Db 2680 CAACCTTCCCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2739  
 QY 866 ArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeuAlaArg 885  
 Db 2740 AGGATCTTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2799  
 QY 886 GlnArgLeuLeuGluArgSerTrpSerValProValIleArgHisLeuPheAlaProLeu 905  
 Db 2800 CAGAACTGCTGGAAGAGCTCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2859  
 QY 906 LysGluTyrPheAlaCys 911  
 Db 2860 AAGGACTACTTTCATGT 2877  
 RESULT 15  
 AF176228 4267 bp mRNA linear PRI 26-OCT-1999  
 LOCUS 4267 bp mRNA linear PRI 26-OCT-1999  
 DEFINITION Homo sapiens DNA cytosine-5 methyltransferase 3B (DNMT3B) mRNA, complete cds.  
 ACCESSION AF176228  
 VERSION AF176228.1 GI:6118091  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 4267)  
 AUTHORS Xu,G.-L., Bestor,T.H., Bourc'hie,D., Haiech,C.-L., Tommerup,N., Bugge,M., Hulten,M., Qu,X., Russo,D.J. and Vilegas-Pequignot,E.  
 TITLE Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene  
 JOURNAL Nature (1999) in press  
 REFERENCE 2 (bases 1 to 4267)  
 AUTHORS Xu,G.-L. and Bestor,T.H.  
 TITLE Direct Substitution  
 JOURNAL Submitted (06-AUG-1999) Genetics and Development, Columbia University, 701 West 168 St., New York, NY 10032, USA  
 FEATURES  
 source  
 1..4267  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="20"  
 /map="20q11-q13"  
 1..4267  
 /gene="DNMT3B"  
 265..2802  
 /gene="DNMT3B"

```

/Note="mutated in human chromosome instability and
immunodeficiency disease known as ICF syndrome; contains
alternative 5' exon 1B"
/codon_start=1
/product="DNA cytosine-5 methyltransferase 3B"
/protein_id="AAF04015.1"
/db_xref="GI:16118092"
/translation="MEPSPEPSSLESKMGKTRHLNGEEDAGGREDSSILVNGACSDSS
DSPLRTEKSPVIRGRSSRLSKREYSSLYTDLTGDDGDDGSDTPVMPKL
PRETRTESPAVRNNNSVSSRRHRSPSTRGRQGRNHDVDESPVFPATSLRR
RATASAGTPWSPSPSYGLTIDTDDTGTQSSSTPYARLAQDSQGGMESQVE
ADSDGSSSYQDGKFGIDLVMGKIKGFSWPMVVSWKATSKRQAMGRWYQWF
DGKFSVADKLVALGLFSGHFNLAENKLVSRKAMVHALEKARVAGKTFPSSPG
DSLEQDKPLMEWAHGGFKPTGIEGLKPNNTQENKTRRATDSDATDCTPAKRLK
TNCYNGKRDDEQDSQREOMASDVANNKSLSDSCGRKNPVSFPHLFGGLCQTC
RDRLELFMYDDGQSYCTVCCGRELKLSNCSRCFCVECLVLTGTAAEA
KLOPWSCYMCLPQCHGVLRKRDWVRLOAFSTSDTGLEAPKLPALPAARRRP
IRLSLFDGIATGYLVKELGTVKGVYVASEVEESIAVGTVEKGNIKYVNDVRNIT
KNIIEGVPEDLVGGSPCNDLSNVNPKARKGLVEGTORLFFEFYHLLNYSRPGKDDR
PFPMFENVAMKVGDKDISRFLECNPMIDAIKVSAAHRYFWGNLPGNRPVIA
SKDKLEQDCLYENRIAKLVQITTKNSIKQGNQLFPVVMNGKEDVLWCTELE
RIFGFPVHYTVDSNMGRGARQKLLGRSWSVIRHLFAPLKYFACE"
BASE COUNT      1069 a 1098 c 1147 g 953 t
ORIGIN

```

```

Alignment Scores:
Pred. No.:      1.37e-79      Length:      4267
Score:          2203.00      Matches:      455
Percent Similarity: 60.75%      Conservative: 110
Best Local Similarity: 48.92%      Mismatches: 221
Query Match:      44.64%      Indels:      145
DB:                9          Gaps:      17

```

US-09-720-086-7 (1-912) x AFI76228 (1-4267)

Db	767	---	CGCCATGGC	---	799
Qy	190	PheGlnAlaGlyAspProTyrTyrIleSerLysArgLysArgAspGluTyrLeuAlaArg	209		
Db	797	TTA	---	799	
Qy	210	TrpLysArgGluAlaGluLysLysAlaLysValIleAlaGlyMetAsnAlaValGluGlu	229		
Db	799	---	---	799	
Qy	230	AsnGlnGlyProGlyGluSerGlnLysValGluGluAla	247		
Db	800	---	CCATCGACCTCACAGACGACACAGAGCACACATGGGAGCGCCAGAGCA	850	
Qy	248	GlnGlnProThrAspProAlaSerProThrValAlaThr	263		
Db	851	GCAGTATCCCTTAGCCCGCTAGCCAGGACGACGAGGGGGCGGATGGAGTCCCGC	910		
Qy	264	ProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAspGluProGlu	283		
Db	911	AG-GTGGAGGCACACAGTGGAGT	948		
Qy	284	TyrGluAspGlyArgGlyPheGlyIleGlyLeuValTrpGlyLysLeuArgGlyPhe	303		
Db	949	TATCAGGATGGGAAGGAGTTTGGATAGGGACCTCGTGTGGGGAAGATCAAGGGCTTC	1008		
Qy	304	SerTrpTrpProGlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGlu	323		
Db	1009	TCCTGTGGCCCGCATGGTGGTGTCTTGAAGGCCACCTCCAAGCGACAGGCTATGTCT	1068		
Qy	324	GlyThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLys	343		
Db	1069	GGCATGCGGTGGTCCAGTGGTGTGGCGATGGCAAGTCTCCAGAGTCTCTGCAGACAAA	1128		
Qy	344	LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnPro	363		
Db	1129	CTGGTGGCACTGGGGCTGTTCAGCCAGCACCTTTAATTTGGCCACCTTCAATAAGCTGTC	1188		
Qy	364	MetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu	383		
Db	1189	TCCTATCGAAAGCCATGTATCATCTCTTGAGAAAGCTAGGCTGCGACTGGCAAGACC	1248		
Qy	384	PheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValGluValGlnAsn	403		
Db	1249	TTCCCC	1287		
Qy	404	LysProMetIleGluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGlu	423		
Db	1288	AAGCCCATGTGTGAGTGGGCCCCACGGGGCTTCAAGCCCACTGGGATCGAGGGCTCTCAA	1347		
Qy	424	ProProGluGluGluLysAsnProTyrLysGluValTyrThrAspMetTrpValGluPro	443		
Db	1348	CCC	1395		
Qy	444	GluAlaAlaAla	461		
Db	1396	GACTCAGCACCTCTGACTACTTCCCGCCACCCAGCGCCCTCAAGACAAATTTGCTATAAC	1455		
Qy	462	AlaGluLysProLysValLysGluIleAspGluArgThrArgGluArgLeuValTyr	481		
Db	1456	AACGGCAAAACCGAGGGATGAA	1506		
Qy	482	GluValArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsn	501		
Db	1507	GATGTTCGCCAACAAACAGAGCAGCCTCGAAGATGGCTGTGTGTCTTGTGGCAGGAAAAAC	1566		
Qy	502	ValThrLeuGluHisProLeuPheValGlyGlyMetCysGlnAsnCysLysAsnCysPhe	521		
Db	1567	CCCGTGTCTTCCACCTCTCTTTGAGGGGGGGCTGTGTGACACATGCGGGATCGCTTC	1626		
Qy	522	LeuGluCysAlaTyrGlnTyrAspAspGlyTyrGlnSerTyrCysThrIleCysCys	541		
Db	1627	CTTGAGCTGTTTTACATGTATGATGACGATGGCTATCAGTCTTACTGCAGCTGTGTGTGCTGC	1686		

BASE COUNT

1069 a

1098 c

1147 g

953 t

US-09-720-086-7

(1-912)

x AF176228

(1-4267)

Alignment Scores:

Pred. No.:

1.37e-79

Length:

4267

Score:

2203.00

Matches:

455

Percent Similarity:

60.75%

Conservative:

110

Best Local Similarity:

48.92%

Mismatches:

221

Query Match:

44.64%

Indels:

145

DB:

9

Gaps:

17

/note="mutated in human chromosome instability and immunodeficiency disease known as ICF syndrome; contains alternative 5' exon 1b"

/codon\_start=1

/product="DNA cytosine-5 methyltransferase 3b"

/protein\_id="AAP04015.1"

/db\_xref="GI:6118092.1"

/translations="MEPSPEPSPLESMMKQDTRHLNGEEDAGGREDSILVNGACSDSSDSPILAEIKRGRSSRLKREVSSILLSYTDLTGDDGGDDGSDTPWPKLPRATRKSPEAVRKNNNVSRHRPSTRGQGRNHNVEDSPVEFATRSRRADSGDSESYQDGKFLGDLGWKLIKGSFWMPVMVSKATSKQAMSGMRWQMFGDGFSVADKLVALGULFQHFNLATPNKLVSIRKAMYHALEKARVAGTKFPSPGDSLEQDKLMLEWAGGFKPTGIEGLKPNNTQENKRRRTADSDATSQCPKRLKTDLYNNGKQDEGSRQMASDVANNKSLSDGSCGRKNPVSFHLPEEGLCQTCRDLFEUFTYDDGYSYTCVCCBGRLLLSNTCCRCFCECLVLTGTAABAIRKQPSYWCMLPQYQGVRRGLLRDMWRVLOAFDFTSDGLEAPKLYPALPAARRRPILNLSFDGLATGVLGLKELGKGVYASEVCEESIAVTGVKHEGNIKKYNVDRNITKNIIEGPEFDLIVIGSCNLSNVNPAKRGLEYGTRFFEHLLNLYSRKPGDDRPFMMFENVAMVKDQKDIISRFLECPNMIDAIKVSAARHVRPWGNIPLGMRPVIASPKMDQVQVLEKNRIKKLQVITITTKNSIKQKQNLFFVVMNGKEDVLMCTELERIFGFPVHYTVDVSNMGRGAKLGRSVFVIRHLFAPLKDYFACE"

Qy 542 GlyGlyValGluValLeuMetCysGlyAsnAsnAsnCysArgCysPheCysValJlu 561  
Db 1687 GAGGCGCCGAGAGCTGCTCTTGGCAGCAACAGAGCTGCTGCGGTGTTCTGTGTGAG 1746  
Qy 562 CysValAspLeuLeuValGlyProGlyValAlaAlaAlaIleGlyGluAspProTrp 581  
Db 1747 TGCCCTGAGGTGCTGTGGGACAGAGCCAGCGCCGAGCCCAAGCTTCAGAGCCCTGG 1806  
Qy 582 AsnCysTyrMetCysGlyHisIleGlyThrTyrGlyLeuLeuArgArgGluAspTrp 601  
Db 1807 AGCTGTTACATGTGTCTCTCCGACGCGCTGTGATGCGCTGCGCGCGAAGAGACTGG 1866  
Qy 602 ProSerArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnIlePheAspProProlys 621  
Db 1867 AACGTGCGCTGACAGGCTTCTTCACAGTGACAGCGGCGTTGAATATGAAGCCCAAG 1926  
Qy 622 ValTyrProProValProAlaGluValArgLysProIleArgValLeuSerLeuPheAsp 641  
Db 1927 CTGTACCTGCTCCATCCGACGCGCGAAGCGGCCCATTCGAGTCTGTCAATGTTGAT 1986  
Qy 642 GlyIleAlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIle 661  
Db 1987 GGCAATCGGAGAGGCTACTAGTCTCTCAAAAGATTGGGCAATGAAGTAAGTAACCTC 2046  
Qy 662 AlaSerGluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIle 681  
Db 2047 GCTTCTGAAGTGTGTGAGAGTCCATTCGTGTGACCGTGAGACAGAGGAGAAATATC 2106  
Qy 682 MetTyrValGlyAspValArgSerValThrGlnLysHisIleGlnGluTrpGlyProPhe 701  
Db 2107 AAATAACGTGAACGACGTAGAGAAATCAAAAGAAATATGAATAATGGGCGCCATTT 2166  
Qy 702 AspLeuValIleGlySerProCysAsnAspLeuSerIleValAsnProAlaArgLys 721  
Db 2167 GACTTGATGATTTGGCGAAGCCCATGCAACGATCTCTCAATGTGAATCCAGCCAGAAA 2226  
Qy 722 GlyLeuTyrGlnGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAla 741  
Db 2227 GGCTGTATGAGGGGTACAGGCGGCTCTTCTCGAATTTTACCACTGCTGAATTACTCA 2286  
Qy 742 ArgProLysGluGlyAspAspArgProPhePheTrpLeuPheGluAsnValAlaIleMet 761  
Db 2287 CGCCCAAGAGGGGTGATGACCGGCGCTTCTGTGATGTTGAATGTTGTAGCCATG 2346  
Qy 762 GlyValSerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAsp 781  
Db 2347 AAGTTGGCGACAGAGGACATCTCACGTTCTCGAGTGTATCCAGTGTATGATTGAT 2406  
Qy 782 AlaLysGluValSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMet 801  
Db 2407 GCCATCAAAAGTTTCTGCTGCTCACAGGCGCCGATCTTCTGGGCACTTACCCGGGATG 2466  
Qy 802 AsnArgProLeuAlaSerThrValAsnAspLysLeuGlnIleCysLeuGlnHis 821  
Db 2467 AACAGGCGCGTATGATCAAAAGATGATAAATCGAGCTCAGAGACTGCTTGAATATC 2526  
Qy 822 GlyArgIleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLys 841  
Db 2527 AATAGGATAGCCAAAGTTAAAGAAATACAGAAATACCAACAGTCGATCAAA 2586  
Qy 842 GlnGlyLysAspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTrpCys 861  
Db 2587 CAGGGGAAACCAACTTTTCCCTGTGTGATGAATGGCAAGAAAGATGTTTGTGTGTC 2646  
Qy 862 ThrGluMetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSer 881  
Db 2647 ACTGAGCTCGAAAGATCTTGTGCTTCTGTGCACTACACAGACTGTGCCAATGGGC 2706  
Qy 882 ArgLeuAlaArgGluArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeu 901  
Db 2707 CGTGTGCCCGCAGAGAGCTGTGGAAAGTCTGTGAGGTGCTGTGCATCGACACCTTC 2766

Qy 902 PheAlaProLeuLysGluTyrPheAlaCys 911  
Db 2767 TTCGCCCTCTGAAGAGACTACTTTCATGT 2796

Search completed: November 22, 2002, 10:57:33  
Job time : 5471.49 secs

**THIS PAGE BLANK (USPTO)**

28	40.2	1.0	635	5	PCT-US94-05354-35	Sequence 35, Appl
29	40.2	1.0	1100	3	US-09-248-335-53	Sequence 53, Appl
30	40.2	1.0	1466	4	US-08-984-919A-10	Sequence 10, Appl
31	40.2	1.0	1466	4	US-08-984-919A-12	Sequence 12, Appl
C 32	40.2	1.0	1472	4	US-08-781-420-10	Sequence 10, Appl
C 33	40.2	1.0	1472	4	US-08-781-420-12	Sequence 12, Appl
34	40.2	1.0	1472	4	US-08-874-102-10	Sequence 10, Appl
C 35	40.2	1.0	1472	4	US-08-874-102-12	Sequence 12, Appl
36	40.2	1.0	1472	4	US-09-006-595A-10	Sequence 10, Appl
C 37	40.2	1.0	1472	4	US-09-006-595A-12	Sequence 12, Appl
38	40.2	1.0	1493	6	5340934-5	Patent No. 5340934
39	40.2	1.0	1875	4	US-08-984-919A-46	Sequence 46, Appl
C 40	40.2	1.0	1875	4	US-08-984-919A-48	Sequence 48, Appl
41	40.2	1.0	1881	4	US-08-874-102-46	Sequence 46, Appl
C 42	40.2	1.0	1881	4	US-08-874-102-48	Sequence 48, Appl
43	40	1.0	2852	3	US-09-027-137-2	Sequence 2, Appl
44	40	1.0	2852	4	US-09-344-441-2	Sequence 2, Appl
45	39.8	0.9	337	2	US-09-032-684-11	Sequence 11, Appl

## ALIGNMENTS

RESULT 1  
US-09-276-531-47  
; Sequence 47, Application US/09276531  
; Patent No. 6183968  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Lal, Preeti  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Yue, Henry  
; APPLICANT: Reddy, Roopa  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Baughn, Mariah R.  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
; TITLE OF INVENTION: RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION  
; NUMBER OF SEQUENCES: 134  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/276,531  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/079,677  
; FILING DATE: March 27, 1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lynn E. Murry, Ph.D.  
; REGISTRATION NUMBER: 42,918  
; REFERENCE/DOCKET NUMBER: PA-0008 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 47:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2077 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: TESTTUT02

THIS PAGE BLANK (USPTO)

CLONE: 12/1435  
US-09-276-531-47

13.1% Score 548.4; DB 4; Length 2077;  
Query Match Similarity 71.4%; Ident. No. 1; Hu-144;  
Beut Local 888; Conservative 0; Mismatches 306; Indels 49; Gaps 11;

2295 GGCCTCTTCTGAGTTTACCACTTCTGAAATATACCCGCCCCCAAGAGAGGAGACAC 2354  
DB 65 GGCCTCTTCTGAAATTTTCCACTCTGAAATTAATACAGCCCAAGAGAGGAGAGAC 123  
OY 2355 GTCATATCTTCTGAGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 2414  
DB 124 GAGCGTTCTTCTGAGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 183  
OY 2415 TCTCAGATTTCTGAGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 2474  
DB TCTCAGATTTCTGAGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 243  
OY 244 ACAGGCGCGGTAATCTTCTGAGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCA 2534  
DB 244 ACAGGCGCGGTAATCTTCTGAGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCA 303  
OY 2535 AGAATGATTAAGCTGAGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 2594  
DB 304 AGAATGATTAAGCTGAGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 363  
OY 2595 AAGTGAAGATTAAGCTGAGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 2654  
DB 364 AAGTGAAGATTAAGCTGAGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 423  
OY 2655 CTGAGTCAATGATGAGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 2714  
DB 424 CTGAGTCAATGATGAGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 483  
OY 2715 GCTTCCCTGCTCACTACAGCAAGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCA 2774  
DB 484 GCTTCCCTGCTCACTACAGCAAGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCA 543  
OY 2775 TGGGAGGTCCTGAGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 2834  
DB 544 TGGGAGGTCCTGAGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 603  
OY 2835 TTGCTGATGATGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 2887  
DB 604 TTGCTGATGATGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 663  
OY 664 CCAGAGATGATGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 2947  
DB 2948 GGCCTCAATCACTGATGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 3007  
OY 722 GTCAT---ACGCTGATGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 777  
DB 3008 TTGAGAGGAGAGCC---GGTGTCTCTCTGAGATGTTGAGAAATGTTGTCGTCATGAAG 3065  
OY 778 TTGAGAGGAGAGCC---GGTGTCTCTCTGAGATGTTGAGAAATGTTGTCGTCATGAAG 837  
DB 3066 TACCCGAGATGATGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 897  
OY 838 CAGCTTAACAGATGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 957  
DB 3126 AAGATGAG---TTTCTTTTACCTGATGATGTTGAGAAATGTTGTCGTCATGAAGTGAAC 3180  
OY 898 AAGATGAG---TTTCTTTTACCTGATGATGTTGAGAAATGTTGTCGTCATGAAGTGAAC 957  
DB 3181 ACCAATAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 3240  
OY 958 ACCAATAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1017  
DB 3241 GATCTAGATGATGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 3300

1018 G-----ACAGTATTTGCAAGAGATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1064  
OY 3301 CATTTGCTGAGATGATGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 3359  
DB 1065 TCTTTACAGAGATGATGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 1124  
OY 3361 AAAAAATCCCT 3420  
DB 1125 TCTTTTCTTCCAGAGGCAAGCAAG---AGAAATGTTGATGATGTTGAGAAATGTTG 1176  
OY 3421 TGGGAGCTTCT 3478  
DB 1177 CCGACATTTCT 1236  
OY 3474 CATTTCCAGATGATGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 3536  
DB 1237 TTTTCCAGATGATGATGTTGAGAAATGTTGTCGTCATGAAGTGAACAGCAAGACA 1279

RESULT 2  
163-14/2

THIS PAGE BLANK (USPTO)



Sequence 1	App1
Sequence 2	App1
Sequence 3	App1
Sequence 4	App1
Sequence 5	App1
Sequence 6	App1
Sequence 7	App1
Sequence 8	App1
Sequence 9	App1
Sequence 10	App1
Sequence 11	App1
Sequence 12	App1
Sequence 13	App1
Sequence 14	App1
Sequence 15	App1
Sequence 16	App1
Sequence 17	App1
Sequence 18	App1
Sequence 19	App1
Sequence 20	App1
Sequence 21	App1
Sequence 22	App1
Sequence 23	App1
Sequence 24	App1
Sequence 25	App1
Sequence 26	App1
Sequence 27	App1
Sequence 28	App1
Sequence 29	App1
Sequence 30	App1
Sequence 31	App1
Sequence 32	App1
Sequence 33	App1
Sequence 34	App1
Sequence 35	App1
Sequence 36	App1
Sequence 37	App1
Sequence 38	App1
Sequence 39	App1
Sequence 40	App1
Sequence 41	App1
Sequence 42	App1
Sequence 43	App1
Sequence 44	App1
Sequence 45	App1
Sequence 46	App1
Sequence 47	App1
Sequence 48	App1
Sequence 49	App1
Sequence 50	App1
Sequence 51	App1
Sequence 52	App1
Sequence 53	App1
Sequence 54	App1
Sequence 55	App1
Sequence 56	App1
Sequence 57	App1
Sequence 58	App1
Sequence 59	App1
Sequence 60	App1
Sequence 61	App1
Sequence 62	App1
Sequence 63	App1
Sequence 64	App1
Sequence 65	App1
Sequence 66	App1
Sequence 67	App1
Sequence 68	App1
Sequence 69	App1
Sequence 70	App1
Sequence 71	App1
Sequence 72	App1
Sequence 73	App1
Sequence 74	App1
Sequence 75	App1
Sequence 76	App1
Sequence 77	App1
Sequence 78	App1
Sequence 79	App1
Sequence 80	App1
Sequence 81	App1
Sequence 82	App1
Sequence 83	App1
Sequence 84	App1
Sequence 85	App1
Sequence 86	App1
Sequence 87	App1
Sequence 88	App1
Sequence 89	App1
Sequence 90	App1
Sequence 91	App1
Sequence 92	App1
Sequence 93	App1
Sequence 94	App1
Sequence 95	App1
Sequence 96	App1
Sequence 97	App1
Sequence 98	App1
Sequence 99	App1
Sequence 100	App1

## ALIGNMENTS

RESULT 1  
US-09-276-531-47  
Sequence 47, Application US/09276531  
Patent No. 6183968  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Yue, Henry  
APPLICANT: Reddy, Roopa  
APPLICANT: Guejler, Karl J.  
APPLICANT: Baughn, Marian R.  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF  
TITLE OF INVENTION-----RECEPTORS AND PROTEINS ASSOCIATED  
NUMBER OF SPOUNCES: 134  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276,531  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/079,677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

```

? NAME: Lynn E. Murry, Ph.D
? REGISTRATION NUMBER: 42,991
? REFERENCE/DOCKET NUMBER: F
? TELECOMMUNICATION INFORMATION
? TELEPHONE: (650) 855-0555
? TELEFAX: (650) 845-4166
? INFORMATION FOR SEQ ID NO: 47
?
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2077 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? IMMEDIATE SOURCE:
? LIBRARY: TESTTUT02
? CLONE: 1271435

```

Alignment Scores:

Pred. NO.:	3,998-52	Length:	207
Score:	746.00	Matches:	151
Percent Similarity:	81.95%	Conservative:	17
Best Local Similarity:	73.66%	Mismatches:	36
Query Match:	15.24%	Indels:	4
DB:	4	Gaps:	1

US-09-720-086-5 (1-908) X US-09-276-531-47 (1-20777)

QY 703 GlySerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeuTyrcGlucIy 722

3 GCGGCNANJGCAACGATCINNAAAT-GTGAATCA-GCCAGGAAGGCTGTATGAGGGAC 60

21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 10

.....

100  
90  
80  
70  
60  
50  
40  
30  
20  
10  
0

263 ArcAcrAcrBhol and ArcAcrBhol-1 mutants

Db 177 GGGACATCTCAGGTTCTGAGGTATTCCACTCATTCATCCCAACCTTTC

QY 783 AlaAlahisArqAlaArqTyrPheTProGlvAsnLeuProGlvMetAsnArqProLeuAla 802

Db 237 GCTGCTCACAGGGCCCGATACCTTCTGGGGCAACCTACCCGGATGAACAGGCCCGTGATA 296

803 SerThrValAsnAspLysLeuGluLeuGlnGluCysLeuGluHisGlyArgIleAlaLys 822

Db 297 GCATCAAGAATGATAA C TCGNGCTGCAGACTGCTTGAATACAA TAGGATAGCCAAG 356

843 PHESETLYSVALARGINILETHRTNRARGSERASNSERIILEYSGINGLYLSPG 842

22 55' 120055551ACGACACAA1AACCCACCAAGTCGAATTCGATCAACAGGGGAAACCCA 416

[illegible][illegible]

4277

883 Arcy,all[.aIG]vArncSerTncSerV[a]DrcV[a]T aXwll:-T a-Cl:-T)-E-:-

Dbb

537 AAGCTGCTGGGAAGTCAATGAGAAGTAGACTCATCCACACCATTTCCTCCTAACA

OY 903 GLUTyrPheAlaCys 907

Db 597 GACTACTTTGCATGT 611

RESULT 2  
US-08-924-345-1

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2002, 08:17:42 ; Search time 543.737 Seconds

(without alignments)  
1780.325 Million cell updates/sec

Title: US-09-720-086-3

Perfect score: 4293  
Sequence: 1 gccgcgcgcacagggcgcgc.....taaaaaaaaaaaaaaa 4293

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

gged: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

N.Geneseq\_101002:.\*  
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	4283.4	99.8	4416	21	AA237097	DNA encoding de no
2	2870.2	66.9	4192	21	AA237095	DNA encoding de no
3	2752.2	64.1	2938	24	ABL90391	Human polynucleoti
4	764.4	17.8	4145	21	AA237098	DNA encoding de no
5	760.4	17.7	4195	21	AA237096	DNA encoding de no
6	613	14.3	709	24	AB057540	Human colon cancer
7	548.2	12.8	1060	22	AAH33860	Human colon cancer
8	504.6	11.8	622	22	AA100271	Human reproductive
9	492.2	11.5	3017	22	AAH14009	Human CDNA sequenc

10	344.2	8.0	711	23	AA86165	DNA encoding novel
11	324.6	7.6	2077	22	AAE80537	Receptor #25 parti
12	285.2	6.6	301	16	AAT21884	Human gene signatu
13	273	6.4	273	24	ABN97389	Gene #3887 used to
14	273	6.4	273	24	ABL64228	Stomach cancer rel
15	254.6	5.9	546	22	ABA20235	Human nervous syst
16	234.8	5.5	283	23	AA813731	Human nervous syst
17	182.4	4.2	2951	22	AA86166	DNA encoding novel
18	173	4.0	411	22	AA185785	Human polynucleoti
19	171.4	4.0	1705	22	AAE82964	Human DNA methyltr
20	171.4	4.0	3002	21	AAE77455	Human ORF3010
21	126.2	2.9	448	22	ABA43755	Human breast cell
22	126.2	2.9	448	22	ABA54212	Human foetal liver
23	126.2	2.9	448	22	ABA23961	Probe #2427 for ge
24	126.2	2.9	448	22	AAK02490	Human brain expres
25	126.2	2.9	448	22	AAK27928	Human bone marrow
26	126.2	2.9	448	22	AA113506	Probe #2439 for ge
27	126.2	2.9	448	22	AA133860	Probe #2546 used t
28	126.2	2.9	448	22	AA102414	Probe #2405 used t
29	126.2	2.9	448	24	ABS02382	Human genome-deriv
30	107.4	2.5	762	22	AAH05250	Human CDNA clone (
31	86	2.0	237	24	ABL77340	Human ovarian canc
32	86	2.0	573	24	ABL82861	Human ovarian canc
33	78.2	1.8	430	22	AA185778	Human polynucleoti
34	75.2	1.8	456	22	AA180464	Human polynucleoti
35	69.6	1.6	326	23	ABV61126	Human prostate exp
36	69.2	1.6	585	23	ABV58085	Human prostate exp
37	68.2	1.6	10732	21	AAH10594	Gene encoding a su
38	66.8	1.6	7434	24	AAH28386	Human chemically t
39	66	1.5	6486	24	ABQ67050	Human angiogenesis
40	65.8	1.5	40662	24	ABL34072	Human immune syste
41	65.4	1.5	14615	22	AA546704	Tumour suppressor
42	65.4	1.5	14798	24	ABL33032	Human immune syste
43	65.2	1.5	432	23	ABV58007	Human prostate exp
44	64.8	1.5	420	23	ABV58336	Human prostate exp
45	64.6	1.5	6359	24	ABK39945	Human chemically p

#### ALIGNMENTS

RESULT 1	AA237097	standard; DNA; 4416 BP.
ID	AA237097	
AC	AA237097;	
DT	27-MAR-2000	(first entry)
XX		
DE	DNA encoding de novo DNA cytosine methyltransferase DNMT3A.	
XX		
KW	De novo DNA cytosine methyltransferase; DNMT3A; neoplastic disorder;	
KW	carcinoma; sarcoma; leukaemia; DNA methylation; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	353..3091
FT		/tag= a
FT		/product= "de novo DNA cytosine methyltransferase
XX		DNMT3A"
XX		
XX	PN	W0967397-A1.
XX		
PD	29-DEC-1999.	
XX		
PF	25-JUN-1999;	99WO-US14373.
XX		
PR	25-JUN-1998;	98US-0090906.
XX		
PR	24-JUL-1998;	98US-0093993.
XX		
PA	(GEHO ) GEN HOSPITAL CORP.	
XX		



QY 1681 GCAGAAATGCGCGAACAATTAGAGACATCTGCATCTCTGTGGAGACCTCAATGTTACCT 1740  
DB 1804 GCAGAAATGCGCGAACAATTAGAGACATCTGCATCTCTGTGGAGACCTCAATGTTACCT 1863  
QY 1741 GGAACACCCCTCTGTGGAGAAATGTCGCAAACTGCAAGAACTGTTCTGAGATG 1800  
DB 1864 GGAACACCCCTCTGTGGAGAAATGTCGCAAACTGCAAGAACTGTTCTGAGATG 1923  
QY 1801 TGGGTAACAAGACAGACAGCGCTACAGTCTTACCTGACCACTGCTGTGGAGCG 1860  
DB 1924 TGGGTAACAAGACAGACAGCGCTACAGTCTTACCTGACCACTGCTGTGGAGCG 1983  
QY 1861 TGAAGTCTCATGTGCGGAAACAACATGCTGAGGTGCTTTTGTGAGTGTGGA 1920  
DB 1984 TGAAGTCTCATGTGCGGAAACAACATGCTGAGGTGCTTTTGTGAGTGTGGA 2043  
QY 1921 CCTCTTGTGGGCGCGGGGCTGCGGAGGACCTTAAGAAAGACCTCTGAACTGCTA 1980  
DB 2044 CCTCTTGTGGGCGCGGGGCTGCGGAGGACCTTAAGAAAGACCTCTGAACTGCTA 2103  
QY 1981 CATGTGCGGACACAAGGTAACCTACGGGCTGTGCGGCGGAGAGGACTGGCCCTCCG 2040  
DB 2104 CATGTGCGGACACAAGGTAACCTACGGGCTGTGCGGCGGAGAGGACTGGCCCTCCG 2163  
QY 2041 GGTCCAGATGTTCTGTCTAATAACAACAACAAGAAATTTGACCTCTCAAAAGTTTACC 2100  
DB 2164 GGTCCAGATGTTCTGTCTAATAACAACAACAAGAAATTTGACCTCTCAAAAGTTTACC 2223  
QY 2101 ACCTGTCCCAAGCTGAGAAAGAAAGCCCATCCGGGTCTGTCTCTTTGATGGAATGCG 2160  
DB 2224 ACCTGTCCCAAGCTGAGAAAGAAAGCCCATCCGGGTCTGTCTCTTTGATGGAATGCG 2283  
QY 2161 TACAGGAGCTCTGTGTGTAAGACTTGGGCAATTAGGTGGAACCGCTACATTTGCTGGA 2220  
DB 2284 TACAGGAGCTCTGTGTGTAAGACTTGGGCAATTAGGTGGAACCGCTACATTTGCTGGA 2243  
QY 2221 GGTGTGAGAGACTCATCAACGCTGAGCATGTGCGGACACAGGAGAAATCAATGTAAGT 2280  
DB 2344 GGTGTGAGAGACTCATCAACGCTGAGCATGTGCGGACACAGGAGAAATCAATGTAAGT 2403  
QY 2281 CGGGGACGTCCGAGAGCTCAACAAGAACATTCAGAGAGTGGGCGCATTCGATCTGCT 2340  
DB 2404 CGGGGACGTCCGAGAGCTCAACAAGAACATTCAGAGAGTGGGCGCATTCGATCTGCT 2463  
QY 2341 GATTTGGGGGAGTCCCTGCAATGACCTCTCATCTGCAACCTGCTGGCAAGGCGCTCTA 2400  
DB 2464 GATTTGGGGGAGTCCCTGCAATGACCTCTCATCTGCAACCTGCTGGCAAGGCGCTCTA 2523  
QY 2401 CGAGGAGCACTGGCCGGCTCTTTCTTGAATCTACCGGCTCTGATGATGCGGGCCCAA 2460  
DB 2524 CGAGGAGCACTGGCCGGCTCTTTCTTGAATCTACCGGCTCTGATGATGCGGGCCCAA 2583  
QY 2461 GAGAGGAGATGATCGCCCTCTTCTTGGCTCTTTGAGAAATGTGTGGCAATGGGCGTTAG 2520  
DB 2584 GAGAGGAGATGATCGCCCTCTTCTTGGCTCTTTGAGAAATGTGTGGCAATGGGCGTTAG 2643  
QY 2521 TGAACAAGAGGAAATCTCGGATTTCTGAGTCCAACCTGTGATGATTTGATGCAAGA 2580  
DB 2644 TGAACAAGAGGAAATCTCGGATTTCTGAGTCCAACCTGTGATGATTTGATGCAAGA 2703  
QY 2581 AGTGTACGTGACACACAGGCGCGCTACTCTTGGGGTAACTTCCGGTATGAAACAGGCC 2640  
DB 2704 AGTGTACGTGACACACAGGCGCGCTACTCTTGGGGTAACTTCCGGTATGAAACAGGCC 2763  
QY 2641 GTTGGCATCACTGTGATGATTAAGCTGAGCTGACAGAGTGTCTGAGCAATGGCAGAT 2700  
DB 2764 GTTGGCATCACTGTGATGATTAAGCTGAGCTGACAGAGTGTCTGAGCAATGGCAGAT 2823  
QY 2701 AGCCAAATTCAGCAAAATGAGAGCAATTAATGAGAGTCAAACTCTTAAGCAGGGCAA 2760  
DB 2824 AGCCAAATTCAGCAAAATGAGAGCAATTAATGAGAGTCAAACTCTTAAGCAGGGCAA 2883

QY 2761 AGACACGATTTTCTGTCTTCTGATGATGAGAAAGAGACATTTATGCTGACCTGAAT 2820  
DB 2884 AGACACGATTTTCTGTCTTCTGATGATGAGAAAGAGACATTTATGCTGACCTGAAT 2943  
QY 2821 GGAAGAGATTTGTTCCAGTCCACTATATGACGTCTCCAAATAGAGCGCTTGGC 2880  
DB 2944 GGAAGAGATTTGTTCCAGTCCACTATATGACGTCTCCAAATAGAGCGCTTGGC 3003  
QY 2881 GAGGACAGACTGCTGGCGGCTCATGAGCGTCCAGTATCCGCACTTCTGCTCC 2940  
DB 3004 GAGGACAGACTGCTGGCGGCTCATGAGCGTCCAGTATCCGCACTTCTGCTCC 3063  
QY 2941 GCTGAAGAGATTTTGGCTGTGTGTAAGGACATGAGGCAAACTGAGTACGACACA 3000  
DB 3064 GCTGAAGAGATTTTGGCTGTGTGTAAGGACATGAGGCAAACTGAGTACGACACA 3123  
QY 3001 AAGTTAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 3060  
DB 3124 AAGTTAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 3183  
QY 3061 AGAAGATATCAGACACCCAGAGAGAGAAATTTAAACAAACAAACAGAGCGGA 3120  
DB 3184 AGAAGATATCAGACACCCAGAGAGAGAAATTTAAACAAACAAACAGAGCGGA 3243  
QY 3121 AATACCGAGGCTTGTGCTGCAAGAGGTTGACATCATCTGATTTTCAATGT 3180  
DB 3244 AATACCGAGGCTTGTGCTGCAAGAGGTTGACATCATCTGATTTTCAATGT 3303  
QY 3181 TATTTCTTCACTCTATTTAAACAAACAGCTCCCTTCTCTCTCCCTTCCCTT 3240  
DB 3304 TATTTCTTCACTCTATTTAAACAAACAGCTCCCTTCTCTCTCCCTTCCCTT 3363  
QY 3241 TTTTTCGCTCAGACCTTTATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 3300  
DB 3364 TTTTTCGCTCAGACCTTTATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 3423  
QY 3301 TTTTGTCTTCTGCTGACGAAACAAGAGTTATTGACGCAAAATCAGTAACAAA 3360  
DB 3424 TTTTGTCTTCTGCTGACGAAACAAGAGTTATTGACGCAAAATCAGTAACAAA 3483  
QY 3361 ATGTAAACAATACCTTTCAGAGAGAAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420  
DB 3484 ATGTAAACAATACCTTTCAGAGAGAAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 3543  
QY 3421 GAAATCATATATGAGGTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3480  
DB 3544 GAAATCATATATGAGGTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3603  
QY 3481 TTTTACTATATATCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 3540  
DB 3604 TTTTACTATATATCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 3663  
QY 3541 GGGACGAAAGAGAGACACTCAGGCGGACGATTTCCCTCCAGCACTGAGCTGTGCT 3600  
DB 3664 GGGACGAAAGAGAGACACTCAGGCGGACGATTTCCCTCCAGCACTGAGCTGTGCT 3723  
QY 3601 GCCAGACCATTTCTGTGACGCAAAACAGAACCCGATTAGCAGCGAGAGAGAGAGACA 3660  
DB 3724 GCCAGACCATTTCTGTGACGCAAAACAGAACCCGATTAGCAGCGAGAGAGAGACA 3783  
QY 3661 CCACACAAGACATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3720  
DB 3784 CCACACAAGACATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3843  
QY 3721 AATCACTTCTAGAACCGCTGTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3780  
DB 3844 AATCACTTCTAGAACCGCTGTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3903  
QY 3781 TGAATATATATTAAGAGTACTGTTAACTATGTAACAACCGACTTCAATATGCTGCT 3840  
DB 3904 TGAATATATATTAAGAGTACTGTTAACTATGTAACAACCGACTTCAATATGCTGCT 3963  
QY 3841 TTCAAACAGGAGATGATTAACAAACATCAGCTTCCAGTTGCTTGGCAAGGGTTT 3900

```

Db 3964 TTCAAAACAGGAGATGAGTAAACATCAGCTTCCAGTTGCTTCTGCGCAAGGGTTT 4023
Qy 3901 CACCAAGATGGAGAAAGGAGACAGCTTGCAGATGGCGGTTTCTCACGGTGGGCTTTC 3960
Db 4024 CACCAAGATGGAGAAAGGAGACAGCTTGCAGATGGCGGTTTCTCACGGTGGGCTTTC 4083
Qy 3961 CCCTTGGTTTGTAAACCAAGTGAAGGAGAGAACTTGGAGCCAGGTTCTCCCTGCCAAA 4020
Db 4084 CCCTTGGTTTGTAAACCAAGTGAAGGAGAGAACTTGGAGCCAGGTTCTCCCTGCCAAA 4143
Qy 4021 AGGGGCTAGATGAGTGTGGCGCCCTGGACAGCTGAGTGGGATTCATCCAGACTC 4080
Db 4144 AGGGGCTAGATGAGTGTGGCGCCCTGGACAGCTGAGTGGGATTCATCCAGACTC 4203
Qy 4081 ATGCATTAACCTTTGATTTGTTTCTTAAGGAGACTCTCTGGCAAGATGCGAGGTT 4140
Db 4204 ATGCATTAACCTTTGATTTGTTTCTTAAGGAGACTCTCTGGCAAGATGCGAGGTT 4263
Qy 4141 ACGGAGTCTTCAAGCCAGTTTCTCACTTTTAGCCAAATTCAGGGCTCTCTTGTGGGAT 4200
Db 4264 ACGGAGTCTTCAAGCCAGTTTCTCACTTTTAGCCAAATTCAGGGCTCTCTTGTGGGAT 4323
Qy 4201 CAGAACTTAATCCAGAGTGTGGAAAGTGACAGTCAAAACCCCACTGGAGCAAAATAAAA 4260
Db 4324 CAGAACTTAATCCAGAGTGTGGAAAGTGACAGTCAAAACCCCACTGGAGCAAAATAAAA 4383
Qy 4261 AACATCAAAACCTTAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4293
Db 4384 AACATCAAAACCTTAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4416

```

## RESULT 2

```

ID AAZ37095
XX AAZ37095 standard; DNA; 4192 BP.
AC AAZ37095;
XX
DT 27-MAR-2000 (first entry)
DE DNA encoding de novo DNA cytosine methyltransferase Dnmt3a.
KW De novo DNA cytosine methyltransferase; Dnmt3a; neoplastic disorder;
KW carcinoma; sarcoma; leukaemia; DNA methylation; ss.
XX
OS Mus sp.

```

```

XX Key Location/Qualifiers
FH 217..2943
FT CDS /*tag= a
FT /product= "de novo DNA cytosine methyltransferase
FT Dnmt3a"

```

```

XX W09967397-A1.
XX
XX 29-DEC-1999.
XX
XX 25-JUN-1999; 99WO-US14373.
XX
XX 25-JUN-1998; 98US-0090906.
XX 24-JUL-1998; 98US-0093993.
XX
XX (GEO ) GEN HOSPITAL CORP.
XX
XX Li E, Okano M, Xie S;
XX
XX WPI; 2000-106298/09.
XX P-PSDB; AAY54055.
XX
XX New mouse and human polypeptides, useful to treat and diagnose
XX neoplastic disorders e.g. carcinomas, sarcomas and leukemias -
XX Claim 8; Fig 1A; 114pp; English.
XX
PS

```

```

XX The present sequence encodes a murine de novo DNA cytosine
CC methyltransferase designated Dnmt3a. The polypeptides can be administered
CC therapeutically, especially by expressing encoding polynucleotides, to
CC treat diseases associated with DNA cytosine methyltransferase, such as
CC neoplastic disorders e.g. carcinomas, sarcomas and leukemias. They can
CC be used to diagnose, or determine susceptibility to neoplastic disorders,
CC by assaying for polypeptide expression levels in mammalian cells/body
CC fluids. They are useful to screen for compounds inhibiting/activating the
CC polypeptide. The polypeptides can also be used in vitro de novo or
CC methylation of DNA. Such in vitro methylation may be used to direct or
CC regulate DNA expression in biological systems, e.g. recombinant DNA
CC methylated in vitro may be introduced into a cell/organism to increase
CC or decrease expression of a desired polypeptide for which the native DNA
CC is under-methylated or not methylated. The polypeptides can also be
CC used to produce antibodies which are useful to detect and purify the
CC polypeptide or therapeutically e.g. to treat neoplastic disorders. The
CC polynucleotides are useful to produce probes and primers which are
CC useful diagnostically.

```

XX SQ Sequence 4192 BP; 1095 A; 1070 C; 1182 G; 844 T; 1 other;

Query Match 66.9%; Score 2870.2; DB 21; Length 4192;  
 Best Local Similarity 83.7%; Pred. No. 0;  
 Matches 3573; Conservative 0; Mismatches 584; Indels 113; Gaps 24;

```

Qy 24 CGGGCGGCGCGGACCCACCGGCGCATACGGTGGAGCCATCGAAGCCCCACCCACAGGCT 83
Db 17 CGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 76
Qy 84 GACAGAGGCGCGGTTTCAACAGAGGGGCTCAACCGGGATCTATGTTAAAGTTTAACTCT 143
Db 77 G-----CACTTTTCGAGGGCTTGACATCAGGCTCTATGTTAAAGTTTAACTCT 126
Qy 144 CGCTCCAAAGACCAACGATATTCCTTTCCCAAGCCG-AGCAGCCCCCAGC-CCCGCG 201
Db 127 TGGTTAAGAACCCACCGGCAATTCCTCTCTGAGCCCTCGACGCCCAACAGCCCTCG 186
Qy 202 CAGCCCGCGCTGCTCGCGCGCCAGATGCGCCGCTGCGCTCGCGCGCGCGCGGGA 261
Db 187 CAGCCCGCGCTGCTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 239
Qy 262 CACCAAGAGCTCTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 321
Db 240 CACCAAGAGCTCTCTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 299
Qy 322 GCGCGTGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 381
Db 300 GAACCGTGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 359
Qy 382 TGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 441
Db 360 TGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 419
Qy 442 GATCTCCAAGTCCCATGCGCCAGGACTCAGGCGCTCAGAGCTATTACCAATGG 501
Db 420 GACCAAGGCTTCAGGCCCATGCGCCAGGACTCAGGCGCTCAGATCTGCTACCAATGG 479
Qy 502 GGACTTGGAGAGCGGAGTGAAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 561
Db 480 AGACTTGGAGAGCGGAGTGAAGCCCAACCTCAGGAGGAGGAGGAGGAGGAGGAGGAGG 539
Qy 562 GGGCGGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 621
Db 540 GGGTGGGCGCGGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 596
Qy 622 AGTGGAAATGGCTGCTGCGACCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 681
Db 597 TGTGGAGAAATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 656
Qy 682 AGAACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 741
Db 657 AGAACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 716

```

QY 742 GGGTGGCTTGGGCTGGGAGTCCAGCTTCCTGCAAGGCTCCATGCCAGGCTCACCTTCCA 801  
DB 717 AGGGGGCTTGGGCTGGGAGTCCAGCTTCCTGCAAGGCTCCATGCCAGGCTCACCTTCCA 776  
QY 802 GGGGGGGAGCCCCCTACTACATCAGCAAGCGGAGCGGGAGCGAGTGGCTGGGAGCCCTGGAA 861  
DB 777 GGCAGGGGAGCCCCCTACTACATCAGCAAGCGGAGCGGGAGCGAGTGGCTGGGAGCCCTGGAA 836  
QY 862 AAGGAGGCTGAGGAAGAAAGCCAAAGTCAATTGACAGAAATGAATCTGTGAAAGAAACA 921  
DB 837 AAGGAGTGTGAGGAAGAAAGCCAAAGTAAATTGACATATGATCTGTGAAAGAAACA 896  
QY 922 GGGGGCCGGGGAGTCTGAGAAAGTGGAGAGAGCCAGCCCTCTCTGTGACAGAGCCAC 981  
DB 897 GGCTCTGAGAGTCTGAGAAAGTGGAGAGAGCCAGCCCTCTCTGTGACAGAGCCAC 956  
QY 982 TGAGCCCCGATCCCCCACTGGGCTACCAAGCTGAGCCGCTGGGGGTCCCATGCTGGGGA 1041  
DB 957 GGAGCCCTGCTTCTCGAGCTGGGCAACCCCTGAGCCAGTAGAGAGGGAGTGGCTGGGGA 1016  
QY 1042 CAAGATATGCCCAAGAGAGGCGATGACGAGCCAGAGTACAGGAGCGGCGGGCTTTGG 1101  
DB 1017 CAAGAAATCTACCAAGAGCCCGAGCGATGAGCTGAGTATGAGAGTGGCCGGGGCTTTGG 1076  
QY 1102 CATTTGGGAGCTGGTGTGGGGGAACTGGCGGGCTTCTCTGTGGCCAGGCGCATTTGT 1161  
DB 1077 CATTTGAGAGCTGGTGTGGGGGAACTTGGGGGTTTCTTGTGTGGCCAGGCGCATTTGT 1136  
QY 1162 GTCTTGTGTGATGACGGGCGGAGAGCCGAGAGCTGAGAGGACCCGCTGGTATGTGTT 1221  
DB 1137 GTCTTGTGTGATGACAGGCGGAGAGCCGAGAGCTGAGAGGACCTGGTGGTCAATGTGTT 1196  
QY 1222 CGGAGAGCGGCAAAATTTCTAGTGTGTGTGTGAGAGTGAATGCCGCTGAGCTGTTTG 1281  
DB 1197 CGGAGATGGCAAGTTCTCAGTGTGTGTGTGTGAGAGTCAATGCCGCTGAGCTTCTTG 1256  
QY 1282 CAGTGCTTCCAGCAGGCGCAGTAAACAACAGAGCCCATGTACCGCAAGGCTTACGA 1341  
DB 1257 CAGTGCTTCCAGCAGGCGCAGTAAACAACAGAGCCCATGTACCGCAAGGCTTACGA 1316  
QY 1342 GGTCTTGAAGTGGCGAGAGCGGCGGGGAGAGTGTTCGGGTGGCCAGAGCGGA 1401  
DB 1317 AGTCTCTCAGAGTGGCGAGAGCGGCGGGAGAGTGTTCAGCTTGCATGACAGTA 1376  
QY 1402 TGAGAGTGAACCTGCAAGGCGGTGAGGTGACAGAAAGGCTTGAATGGGCTT 1461  
DB 1377 TGAAAGTGAACGTGGCAAGGCTGTGAAAGTGAAGCAAGAGATGATGATGGGCTT 1436  
QY 1462 GGGGGGCTTCCAGGCTTCTGGGCTTAAAGGCTTGAAGCGCAGCAAGAGAGAGAAATCC 1521  
DB 1437 CGGTGGCTTCCAGGCTTCTGGGCTTAAAGGCTTGAAGCGCAGCAAGAGAGAGAAATCC 1496  
QY 1522 CTACAAAGAAAGTGTACACGAGCATGTGGGTGAAACTGAGGAGCTGCTTACGACCAAC 1581  
DB 1497 TTACAAAGAAAGTGTACACGAGCATGTGGGTGAAACTGAGGAGCTGCTTACGACCAAC 1556  
QY 1582 TCCAGCAGCAAAAGGCCCCGGAGAGACAGCGGAGAGAGGCCCCAAGGCTTGAAGATTA 1641  
DB 1557 CCGACCGCAGCAAGAAAGCCAGAAAGAGACAAACAGAGAAACTTAAAGTCAAGGAGATCAT 1616  
QY 1642 TGATGAGCGCAGCAAGAGAGCGGCTGTGTACGAGGTGGCGAGAGAGTGGCCGAACTTGA 1701  
DB 1617 TGATGAGCGCAGCAAGAGAGCGGCTGTGTATGAGGTGGCGCAGAAAGTGGCAAAATCGA 1676  
QY 1702 GGAATCTGCACTCTCTGTGGAGAGCTTCAATGTAACTCTGGAGACCCCTTGTGTTGG 1761  
DB 1677 GGAATTTGTATCTCAGTGTGGAGAGCTTCAATGTCACTCTGGAGACCCATTTCTTATGG 1736  
QY 1762 AGGAATGTGCCAAAACTGCAAGAACTGTCTTCTGAGAGTGGCGAGATGACAGTACGAGCA 1821  
DB 1737 AGGCAATGTGCCAAAACTGTAAAGAACTGTCTTCTGAGAGTGGCTTACCAAGATGACAGCA 1796  
QY 1822 CGGCTACAGTCTCTACTGCAACCATCTGCTGTGGGGGCGGTGAGGTCTCATGTGGGAAA 1881  
DB 1797 TGGGTACAGTCTCTATTTGCAACCATCTGCTGTGGGGGCGGTGAAAGTCTCATGTGGGAAA 1856  
QY 1882 CAACAACTGCTGACAGGTCTTTTGGGTGGAGTGTGTGACCTCTTGGTGGGGCCGGGGGC 1941  
DB 1857 CAACAACTGCTGACAGGTCTTTTGGGTGGAGTGTGTGACCTCTTGGTGGGGCCGAGAGC 1916  
QY 1942 TGCCAGGAGCCCATTTAAGAGAGAGCCCTGAAACTGTCTAATGATGGCGGACAAAGGTAC 2001  
DB 1917 TGCTCAGGAGCCCATTTAAGAGAGAGCCCTGAAACTGTCTAATGATGGCGGACAAAGGTAC 1976  
QY 2002 CTACGGGCTGCTGGCGGCGGAGAGAGCTGGCCCTCCGGGCTCCAGATGTTTCTGCTAA 2061  
DB 1977 CTATGGGCTGCTGGAGAGAGCGGGAAGAGTGGCTTCTGACATCGAATGTTCTTGGCAA 2036  
QY 2062 TAAACAGACCAAGAAATTTGAACCTCCAAAGGTTTACCACCTGTCCAGCTGAGAAAGAG 2121  
DB 2037 TAAACATGACCAAGAAATTTGAACCTCCAAAGGTTTACCACCTGTGCAAGCTGAGAAAGAG 2096  
QY 2122 GAAGCCCATCCGGGTGCTGCTCTCTTGTATGAGATGCTACAGGGGCTCCGTGGCTGAA 2181  
DB 2097 GAAGCCCATCCGGGTGCTGCTCTCTTGTATGAGATGCTACAGGGGCTCCGTGGCTGAA 2156  
QY 2182 GAACTTGGGCAATTCAGGTGAGACCGCTACATTTGCTCGAGGTGTGAGAGATTCATCAC 2241  
DB 2157 GAAGCTGGGCAATTCAGGTGAGACCGCTACATTTGCTCGAGGTGTGAGAGATTCATCAC 2216  
QY 2242 GGTGGGATGTGTGGGACCAAGGGAAGATCATGTACGTGGGGAGCTCCGACAGCTCAC 2301  
DB 2217 GGTGGGATGTGTGGGACCAAGGGAAGATCATGTACGTGGGGAGCTCCGACAGCTCAC 2276  
QY 2302 ACAAGACATATCCAGGAGTGGGGCCCATTCGATCTGGTATTTGGGGGAGTCCCTGCAA 2361  
DB 2277 ACAAGACATATCCAGGAGTGGGGCCCATTCGATCTGGTATTTGGGGGAGTCCCTGCAA 2336  
QY 2362 TGAACCTTCATGCTCAACCTGCTGCGAAGGAGCTTCTACGAGGAGCACTGGCGCTCTT 2421  
DB 2337 TGAACCTTCATGCTCAACCTGCTGCGAAGGAGCTTCTACGAGGAGCACTGGCGCTCTT 2396  
QY 2422 CTTTGAATTTACCGGCTCTGCTGATGATGCGCGGCCCAAGAGGAGATGATGCGCCCTT 2481  
DB 2397 CTTTGAATTTACCGGCTCTGCTGATGATGCGCGGCCCAAGAGGAGATGATGCGCCCTT 2456  
QY 2482 CTTTGGCTCTTTGAGAAATGATGATGCGCGGCTTGAAGTGAAGAGGAGCAATCTGCG 2541  
DB 2457 CTTTGGCTCTTTGAGAAATGATGATGCGCGGCTTGAAGTGAAGAGGAGCAATCTGCG 2516  
QY 2542 ATTTCTGAGTCCAAACCTGTGATGATGATGCGCAAAAGAGTGTCAAGTCCAGCAGGCG 2601  
DB 2517 ATTTCTGAGTCCAAACCTGTGATGATGATGCGCAAAAGAGTGTGTGCTGACACAGGCG 2576  
QY 2602 CCGCTACTTCTGGGATTAACCTTCCGGTATGAAACAGGCGGTGGCATCACTGTGAATGA 2661  
DB 2577 CCGTACTTCTGGGATTAACCTTCCGGTATGAAACAGGCGGTGGCATCACTGTGAATGA 2636  
QY 2662 TAAAGTGAAGTCTGAGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 2721  
DB 2637 TAAAGTGAAGTCTGAGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 2696  
QY 2722 GACCATTAACAGAGTCAAACTCCATTAAGCAGGGGCAAGACAGCATTTTCTGTCTT 2781  
DB 2697 GACCATTAACAGAGTCAAACTCCATTAAGCAGGGGCAAGACAGCATTTTCTGTCTT 2756  
QY 2782 CATGAATGAAGAAAGAGACATCTTATGTGCTCACTGAATGAGAAAGGTATTTGGTTTCC 2841  
DB 2757 CATGAATGAAGAAAGAGACATCTTATGAAGCAGGGGCAAGACAGCATTTTCTGTCTTCC 2816  
QY 2842 AGTCCACTAATACAGTGTCCCAACATGAGCGGCTTGGCGAGGCAAGAGATGCTGGGCGG 2901  
DB 2817 GGTCCACTAACAGAGTGTCCCAACATGAGCGGCTTGGCGAGGCAAGAGATGCTGGGCGG 2876  
QY 2902 GTATGAGAGGCTGCAAGTCAATCCGCAACTTCTGCTCCGCTGAAGAGATTTTGGCTG 2961

Db 2877 ATCTGGAGCGTCCCGGTCTATCCGCCACCTCTTCGCTCCCGCTGAGGAATATTTTGGCTTG 2936  
Qy 2962 TGTGTAAAGGACATGGGGCAAACTGAGGTAGCGA-----CACAAGTTAAACAACAAC 3017  
Db 2937 TGTGTAAAGGACATGGGGCAAACTGAAAGTAGTGATGATATAAAGAGTTTAAACAACAAC 2996  
Qy 3018 -----AAAAAACAACAACATATATAAACAACAAGACATGAGGATGGAGAGATATCA 3072  
Db 2997 AAACAACAAAAACAACAACAATATAAACAACAAGACGAGAGCGGAGAAAGT-TCA 3055  
Qy 3073 GCACCCAGAGAGAGAAAGGAATTTAAACAACAACAACAAGAGCGGAGAAATACCGGAGG 3132  
Db 3056 GCACCCAGAGAGAGAAAGGAATTTAAAGC--AAACAACAAGAGAGAGAAACGCGGAGG 3113  
Qy 3133 CTATTTCGCTTCGAAAGGGTTCGACATCATCTCTGATTTTCAATGTTTATTTCTTCAGTC 3192  
Db 3114 GCTTGGCTTCGAAAGGGTTCGACATCATCTCTGATTTTCAATGTTTAACTTTCAGTC 3173  
Qy 3193 CTATTTAACAACAACAACCAAGCTCCCTTCCCTCTCCCTCTCCCTCTTCTTTTTCGGTCA 3252  
Db 3174 CTATCTAAAGCAAAATAGGC-CCCTCCCTCTCTCCCTCTCCCTCTTAGAGGG 3229  
Qy 3253 GACCTTTATTTCTACTC-TTTTCAGAGGGTTTTCTGTTGTTTGGG-TTTTGTCTCT 3310  
Db 3230 AACTTTGTTGTTTCTACTCTTTTTCAGAGGGTTTTCTGTTGTTTGGGTTTTGTTTCT 3289  
Qy 3311 TGCTGTGACTGAAACAAGAGGTTATTCGACAAAAATCAGTAACAAAAATAGTAACA 3370  
Db 3290 TGCTGTGACTGAAACAAGAGGTTATTCGAGC-AAAAATCAGTAACAAAAATAGTAACA 3348  
Qy 3371 TACCTTTCAGAGAAAGGTGGAGGAGGAGGAAAAAGGAAATTTTAAAGAAATCTATA 3430  
Db 3349 TGCCTTCGAGAGAAAG-----GGAGAGAGGAAATTTCTATAAAACATTAATA 3397  
Qy 3431 TATTGGGTTCTTTTTTTTTTTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTACTA 3490  
Db 3398 TATTGGTTTTTTTTTTTTTTTCTCTT-----TC 3424  
Qy 3491 TATATCTTTTTTTTGTGTTCTTAGCTGATCAGATAGGAGCAACAAGCAGGGGACGGAAA 3550  
Db 3425 TATATATCTCTTTGTTGTTCTAGCTGATCAGATAGGAGCAACAACA--GGAAGAGAA 3482  
Qy 3551 GAGAGACACTCAGCGCGCAGCATTCCTCCAGCCACTGAGCTGCTGTCGACGACCA 3610  
Db 3483 TAGAGACCTCGAGGCGAG-AGTCTCTCTCCACCCCGAGCAGTCTCAACAGCACA 3541  
Qy 3611 TTCTGTGTACGCAAAACAGAACCCAGTTAGCAGCAGGAGAGCAACACCAACACAGA 3670  
Db 3542 TTCCTGTGATGCAAAACAGAACCCAACTAGCAGCGGGCGCTGAGAGAACACACACA 3601  
Qy 3671 CA-TTTTTCACAGTATTTTCAGTCCCTACACACAGGAAACCTTGAAGAAATCAGTTT 3729  
Db 3602 GACACTTTTCACAGTATTTTCAGTCCCTACACACAGGAAACCTTGAAGAAACAGTTT 3661  
Qy 3730 CTAGAAGCCGCTGTTACCTCTTTTACAGTTTATATATATATATATATATATATATAG-ATA 3787  
Db 3662 CTAGAAGCCGCTGTTACCTCTTTTACAGTTTATATATATATATATATATATATATAGATATA 3721  
Qy 3788 TATATATAAAGGTACTGTTAACTACTGTACAAACCCGACTTCTATAATGTTGCTTTTC-AAA 3846  
Db 3722 TATATATAAAGGTACTGTTAACTACTGTACATCCCGACTTCTATAATGTTGCTTTCAAAA 3781  
Qy 3847 CAGCGAGATGAGTAAACAAATCAGCTTCCAGTTGCTTCTGCGCAAGGGTTTTCACCAA 3906  
Db 3782 CAGCGAGATGAGCAAAAGACATCAGCTTCGCGCTCGGCTCTGTGCAAGGGTTTTCAGGCC 3841  
Qy 3907 GGATGGAGAAAGGAGACAGCTTGCAGATGCGCGGTTCTCAGCGTGGGCTCTTCCCTTG 3966  
Db 3842 AGGATGGGAGAGGGGAGCAGCTGGAGGGGGTTTAAACAACATGAAGGATGACCATATC 3901  
Qy 3967 GTTTGTTAACGAAGTGAAGGAGGAGAACTTGGGAGCCAGGTTCTCCCTGCGCAAAAGGGGG 4026

Db 3902 ACCCCCCAC-----CCCTGCCCATGCTAGCTTCACCTGCCCAAAAGGGGC 3948  
Qy 4027 CTAGATGAGTGTGTCGGCCCGTGGACAGCTGAGAGTGGGATTTCATCAGACTCATGCA 4086  
Db 3949 TTAGCTGAGGTGGTCGGACCTGGGAAGCTGAGTGTGGAATTTATCCAGACTCGGTGC 4008  
Qy 4087 TTAACCTTTGATGTTGTTTCTAAAGGAGACTCCCTCGGCAAGATGGCAGAGGGTACGGAG 4146  
Db 4009 ATAAACCTTTAGAAATATGAATCTAAATGACTCCCTCAGAAAAATGGC-----TTGAGAA 4062  
Qy 4147 TCTTCAGGCCAGTTTCTCACTTTAGCCAAATTCGAGGGCTCTTGTGTGGGATCAGAAC 4206  
Db 4063 AACATTTGTCCTGATTTTGAATTCGTGAGCCACGTTGAGGGCCCTTGTGGGATCAGAAA 4122  
Qy 4207 TAATCCAGAGTGTGGGAAAGTGACAGTCAA--AACCCACCTGGAGCAAAATAAAAAACA 4264  
Db 4123 TATTCCAGAGTGGGAAAGTGACCCGCCATTAACCCCTCGGAGCAAAATAAAAAACA 4182  
Qy 4265 TACAAACGT 4274  
Db 4183 TACAAAATGT 4192

## RESULT 3

ABL90391

ID ABL90391 standard; cDNA; 2938 BP.

XX ABL90391;

XX 24-MAY-2002 (first entry)

XX Human polynucleotide SEQ ID NO 953.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
XX antiallergic; hepatotrophic; antidiabetic; antiinflammatory; antitumor;  
XX vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;  
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; human; secreted protein; gene; ss.

XX Homo sapiens.

XX WO200190304-A2.

XX 29-NOV-2001.

XX 18-MAY-2001; 2001WO-US16450.

XX 19-MAY-2000; 2000US-205515P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Birse CE, Rosen CA;

XX WPI; 2002-122018/16.

XX P-PSDB; ABB89982.

XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
XX prevention of neural, immune system, muscular, reproductive,  
XX gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
XX disorders -

XX Claim 4; SEQ ID NO 953; 2081pp + Sequence Listing; English.

XX The invention relates to novel genes (ABL89449-ABL90853) and proteins  
XX (ABB9040-ABB90444) useful for preventing, treating or ameliorating  
XX medical conditions e.g. by protein or gene therapy. The genes are  
XX isolated from a range of human tissues disclosed in the specification.  
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful  
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone  
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune,  
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's



CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 2938 BP; 759 A; 695 C; 774 G; 700 T; 10 other;

Query Match 64.1%; Score 2752.2; DB 24; Length 2938;  
Best Local Similarity 99.1%; Pred. No. 0;

Matches 2761; Conservative 5; Mismatches 16; Indels 4; Gaps 2;

QY 1471 CGAGCCTTCGCGCCCTTAAGGGCCCTGAGAGCCACGAGAAAGAGAAGATCCCTACAAAGA 1530  
9 CGAGCCTTCGCGCCCTTAAGGGCCCTGAGAGCCACGAGAAAGAGAAGATCCCTACAAAGA 68  
DB 1531 AGTGTACAGGACATGTGGGTGGAACCTGAGGAGCTGCTTACGACACCTCCAGC 1590  
69 AGTGTACAGGACATGTGGGTGGAACCTGAGGAGCTGCTTACGACACCTCCAGC 128  
QY 1591 CAAAAGCCCCGGAAGAGGACAGCGGAGAGCCCAAGTCAAGAGATTATTGATGCG 1650  
129 CAAAAGCCCCGGAAGAGGACAGCGGAGAGCCCAAGTCAAGAGATTATTGATGCG 188  
QY 1651 CACAAGAGAGCGCTGGTGTACGAGGTGCGGAGAAAGTGCAGAACTTATAGGACATCTG 1710  
189 CACAAGAGAGCGCTGGTGTACGAGGTGCGGAGAAAGTGCAGAACTTATAGGACATCTG 248  
QY 1711 CATCTCTGTGGAGCCTCAATGTTACCTTGAAACACCCCTCTTCTTGAGAGAAATGTG 1770  
249 CATCTCTGTGGAGCCTCAATGTTACCTTGAAACACCCCTCTTCTTGAGAGAAATGTG 308  
DB 1771 CCAAACTGCAAGAACTGCTTTTCTGAGGTGTGCTACAGTACGACAGCGCTACCA 1830  
309 CCAAACTGCAAGAACTGCTTTTCTGAGGTGTGCTACAGTACGACAGCGCTACCA 368  
QY 1831 GTCTCACTGACCATCTGCTGTGGGGGCGGTGAGGTCTCATGTGGGGAACAAACAATG 1890  
369 GTCTCACTGACCATCTGCTGTGGGGGCGGTGAGGTCTCATGTGGGGAACAAACAATG 428  
QY 1891 CTGCAAGTCTTTTTCGTGTGAGGTGTGGAACCTTTGTGGGGCGCGGGCTGCCAGGC 1950  
429 CTGCAAGTCTTTTTCGTGTGAGGTGTGGAACCTTTGTGGGGCGCGGGCTGCCAGGC 488  
DB 1951 AGCCATTAAAGAAAGACCCCTGGAATCTCTACATGTGCGGGCAAGGGTACCTACGGGCT 2010  
489 AGCCATTAAAGAAAGACCCCTGGAATCTCTACATGTGCGGGCAAGGGTACCTACGGGCT 548  
QY 2011 GCTGCGGCGGAGAGGAGCTGGCCCTCCGGGCTCAGATGTTCTGGCTAATAACACGA 2070  
549 GCTGCGGCGGAGAGGAGCTGGCCCTCCGGGCTCAGATGTTCTGGCTAATAACACGA 608  
QY 2071 CGAGAAATTTGACCTCTCAAAAGTTTACCCACCTGTCCAGCTGAGAAAGAGAGCCCAT 2130  
609 CGAGAAATTTGACCTCTCAAAAGTTTACCCACCTGTCCAGCTGAGAAAGAGAGCCCAT 668  
DB 2131 CGGGGTGCTCTCTCTTTATGTAATGCTACAGGGCTCTCGTGTGTAAGGACTTGGG 2190  
669 CGGGGTGCTCTCTCTTTATGTAATGCTACAGGGCTCTCGTGTGTAAGGACTTGGG 728  
QY 2191 CATTCAAGGTGAGACCGGCAATTCGCTCGAGGTGTGTAAGGACTCCATCAAGGGGACT 2250  
729 CATTCAAGGTGAGACCGGCAATTCGCTCGAGGTGTGTAAGGACTCCATCAAGGGGACT 788  
QY 2251 GGTGCGGACACAGGGAGAGATGATGATCGTGGGAGAGTCCGAGCTCACAGAGCA 2310  
789 GGTGCGGACACAGGGAGAGATGATGATCGTGGGAGAGTCCGAGCTCACAGAGCA 848  
QY 2311 TATTCAGGAAGTGGGCGCCATTCGATCTGTGATTTGGGGGAGTCCCTGCATGACCTTCTC 2370  
|||||

DB 849 TATTCAGGAAGTGGGCGCCATTCGATCTGTGATTTGGGGGAGATCCCTGCATGACCTCTC 908  
QY 2371 CATCGTCAACCCCTGCTCGAAGGGCCCTTACAGAGGCACTGGCCGCTCTTTGAGTT 2430  
909 CATCGTCAACCCCTGCTCGAAGGGCCCTTACAGAGGCACTGGCCGCTCTTTGAGTT 968  
QY 2431 CTACCCGCTCTCTGATGATGCGGCGCCCAAGAGGAGATGATGCGCCCTTCTTGCGCT 2490  
969 CTACCCGCTCTCTGATGATGCGGCGCCCAAGAGGAGATGATGCGCCCTTCTTGCGCT 1028  
DB 2491 CTTTGAAGATGTGGTGGCCATGGGCGTTGATGCAAGAGGAGATCTCGGATTTCTCGA 2550  
1029 CTTTGAAGATGTGGTGGCCATGGGCGTTGATGCAAGAGGAGATCTCGGATTTCTCGA 1088  
QY 2551 GTCCAAACCCCTGTGATGATGATGATCCAAAGAGTGTGAGCTGCACAGAGGCCGCTACTT 2610  
1089 GTCCAAACCCCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1148  
DB 2611 CTGGGGTAACTTTCCGGTATGAAAGGCGGTTGGCATCCACTGTGAATGATTAAGCTGGA 2670  
1149 CTGGGGTAACTTTCCGGTATGAAAGGCGGTTGGCATCCACTGTGAATGATTAAGCTGGA 1208  
QY 2671 GCTGCAAGAGTGTCTGGAGATGCGAGATAGCCAGTTACAGCAAGTGAAGGACATTTAC 2730  
1209 GCTGCAAGAGTGTCTGGAGATGCGAGATAGCCAGTTACAGCAAGTGAAGGACATTTAC 1268  
QY 2731 TACGAGGTCAAACTCCATTAAGAGAGGCAAAAGCAGCAATTTTCTGTCTTATGATGA 2790  
1269 TACGAGGTCAAACTCCATTAAGAGAGGCAAAAGCAGCAATTTTCTGTCTTATGATGA 1328  
DB 2791 GAAAGAGGACATCTTATGTTGCACTGAAATGGAAGGATATTTGGTTCCGATCCACTA 2850  
1329 GAAAGAGGACATCTTATGTTGCACTGAAATGGAAGGATATTTGGTTCCGATCCACTA 1388  
QY 2851 TACTGACGTCTCCAAATGAGCGCTTGGCGAGGAGAGATCTGTGGCCGCTCATAGAG 2910  
1389 TACTGACGTCTCCAAATGAGCGCTTGGCGAGGAGAGATCTGTGGCCGCTCATAGAG 1448  
QY 2911 CGTGCCAGTCACTCGGCACTCTTCTCGCTCGGCTGGAAGGATTTTGGGTGTGTAAG 2970  
1449 CGTGCCAGTCACTCGGCACTCTTCTCGCTCGGCTGGAAGGATTTTGGGTGTGTAAG 1508  
DB 2971 GACATGGGGCAAACTGAGGTAGCGACACAAAGTTAAACAAACAAACAAACAAACAA 3030  
1509 GACATGGGGCAAACTGAGGTAGCGACACAAAGTTAAACAAACAAACAAACAAACAA 1568  
QY 3031 CATTAATTAACCAAGAAATAGAGATGAGAGAAATACAGACCCAGAAAGAAAAA 3090  
1569 CATTAATTAACCAAGAAATAGAGATGAGAGAAATACAGACCCAGAAAGAAAAA 1628  
DB 3091 GGAATTTAAACAAAAACAAGAGCGGAAATACCGAGGGGCTTGGCTTGGAAAAAG 3150  
1629 GGAATTTAAACAAAAACAAGAGCGGAAATACCGAGGGGCTTGGCTTGGAAAAAG 1688  
QY 3151 GTTGAACATCATCTCTGATTTTCAATGTTATCTTCACTATTTAAAAACAAAC 3210  
1689 GTTGAACATCATCTCTGATTTTCAATGTTATCTTCACTATTTAAAAACAAAC 1748  
DB 3211 AAGCTCCCTTCCCTTCTCCCTCTCTTTTTCGATGAGAGCTTTTATTTTCTACT 3270  
1749 AAGCTCCCTTCCCTTCTCCCTCTCTTTTTCGATGAGAGCTTTTATTTTCTACT 1808  
QY 3271 CTTTTCAGAGGGGTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 3330  
1809 CTTTTCAGAGGGGTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1868  
DB 3331 GGTATTTCAGCAAAATAGTAACAAATAATAGTAACAAATACCTTTCAGAGGAGAGGTTG 3390  
1869 GGTATTTCAGCAAAATAGTAACAAATAATAGTAACAAATACCTTTCAGAGGAGAGGTTG 1928  
QY 3391 GGAAGAGAGAAAAAGGAAATTTTAAAGAAATCTATATATTTGGGTTTGTGTTT 3450  
1929 GGAAGAGAGAAAAAGGAAATTTTAAAGAAATCTATATATTTGGGTTTGTGTTT 1985  
|||||

|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|

```

FT XX /number= 23
PV XX MO9696797-AL.
PD XX 29-DEC-1999.
PF XX 25-JUN-1999; 99WO-US14373.
PR XX 25-JUN-1998; 98US-0090906.
PR XX 24-JUL-1998; 98US-0093993.
PA XX (GENO ) GEN HOSPITAL CORP.
PI XX Li E, Okano M, Xie S;
XX WPI; 2000-106298/09.
DR XX P-PSDB; AAY54058.

New mouse and human polypeptides, useful to treat and diagnose
neoplastic disorders e.g. carcinomas, sarcomas and leukemias -

XX
XX Example 1; Fig 1D; 114pp; English.
PS
CC The present sequence encodes a human de novo DNA cytosine
CC methyltransferase designated DNMT3B1. The DNMT3B gene also produces,
CC through alternate splicing, at least two shorter isoforms termed
CC DNMT3B2 (comprising amino acids 1-355 and 376-853 of AAY54058, see
CC AAY54059) and DNMT3B3 (comprising amino acids 1-355, 376-743 and 807-853
CC of AAY54058, see AAY54060) (sequences not given in the specification).
CC The polypeptides can be administered therapeutically, especially by
CC expressing encoding polynucleotides, to treat diseases associated with
CC DNA cytosine methyltransferase, such as neoplastic disorders e.g.
CC carcinomas, sarcomas and leukemias. They can be used to diagnose, or
CC determine susceptibility to neoplastic disorders, by assaying for
CC polypeptide expression levels in mammalian cells/body fluids. They are
CC useful to screen for compounds inhibiting/activating the polypeptide. The
CC polypeptides can also be used for in vitro de novo methylation of DNA.
CC Such in vitro methylation may be used to direct or regulate DNA
CC expression in biological systems, e.g. recombinant DNA methylated in
CC vitro may be introduced into a cell/organism to increase or decrease
CC expression of a desired polypeptide for which the native DNA is
CC under-methylated or not methylated. The polypeptides can also be used to
CC produce antibodies which are useful to detect and purify the polypeptide
CC or therapeutically e.g. to treat neoplastic disorders. The
CC polynucleotides are useful to produce probes and primers which are useful
CC diagnostically.
XX
XX Sequence 4145 BP; 1041 A; 1083 C; 1096 G; 925 T; 0 other;
SQ
Query Match 17.8%; Score 764.4; DB 21; Length 4145;
Best Local Similarity 63.0%; Pred. No. 1e-145;
Matches 1144; Conservative 0; Mismatches 706; Indels 26; Gaps 3

OY 1017 AGCCGCTGGGGTTCGATGCTGGGGACAGAAATCCCAACAAAGCAGGAGCAGGCAG 1076
Db 701 AGCAGGGGGGCGATGAGTCCCGCCAGAGTGGAGCGACAGATGAGATGAGACAGTTCAG 760
OY 1077 AGTACGAGAGACGGCCGGGGCTTTGGCAATGGGGAGCTGCTGGGGGAAAATCGGGGCT 1136
Db 761 AGTATTCAGATGGAGGAGATTGGAAATAGGGGACCTCGTGTGGGAAACATCAAGGGCT 820
OY 1137 TTTCTGTGGGCGCAGGCGCATGTGTCTGTGGATGAGAGGGCCGGAGCGGAGACTG 1196
Db 821 TTTCTGTGGGCGCCGCGCATGTGTCTGTGGATGAGAGGGCCGGAGCGGAGACTG 880
OY 1197 AAGGACCCGCTGGGTCATGTGGTTGAGACGCGCAATTTCTAGTGTGTGTTGAGA 1256
Db 881 CTGGCATGGGGTGGGTCGCAAGTGGTTGGCGATGGCAAGTTCTCGAGGTCTCGACACA 940
OY 1257 AACTGATGGCGCTGACTGTTTGGAGTGCCTTCCACCAAGGCGCAGTACACAGCAGC 1316
Db 941 AACTGTGGCATCTGGGCTGTTTCAGCGACGACATTTAATTGGCCACCTTCAATTAAGCTCG 1000

```

[illegible]

```
Db 2080 AATGTGAATCCAGCCAGGAAGCCCTGTATGAGGGTACAGCCGGCTCTCTTCGAATTT 2139
Qy 2432 TACGGCTCTGTCATGATCGCGGCCCAAGAGGAGATGATCGCCCTCTCTTCTGGCTC 2491
Db 2140 TACCACCTGCTGAATTAATCTACAGCCCAAGAGGGGTGATACCGCCGCTCTCTTCTGGATG 2199
Qy 2492 TTTGAGAAATGTGTGGCCATGCGGCTTTAGTGACAAGAGGGACATCTCGCGATTTCTCGAG 2551
Db 2200 TTTGAGAAATGTTAGCCATGAAGCTTGGCGACAAGAGGGACATCTCAAGGTTCTCGAG 2259
Qy 2552 TCCAAACCTGTGATGATTCATGCCAAAGAGTGTGAGTGCACACAGGCGCCGCTACTTC 2611
Db 2260 TGTAAATCCAGTGATGATGCCATCAAGTCTCTGCTCTCACAGGCGCCGATACTTC 2319
Qy 2612 TGGGTAACCTTCCCGTATGACAGGCGGTTGGATCCACTGTGAATGATGAAGCTGGAG 2671
Db 2320 TGGGCAACCTTACCCGGATGAACAGGCGGCTGTAGATCAAGATGATGAATCGAG 2379
Qy 2672 CTGAGAGTGTCTGGAGCATGGCAGGATACCCAGTTCAGAACTGAGGACCACTACT 2731
Db 2380 CTGAGAGTGTCTGGATACAAATAGGATAGCCAAAGTTAAAGAAATGACAGCAATAACC 2439
Qy 2732 ACGAGGTCAAACTCCATAAGCAGGCGCAAGCCAGCATTTCTCTCTTCATGAATGAG 2791
Db 2440 ACCAAGTCGAATCGATCAACAGGCGGAAACCAACTTTTCCCTGTCTCATGAATGGC 2499
Qy 2792 AAAGAGGACATCTTATGTGTGCACTGAAATGGAAGGGTATTTGGTTTCCAGTCCACTAT 2851
Db 2500 AAAGAAATGTTTGTGTGCACTGAGCTCGAAAGGATCTTTGGCTTTCTCTGCACTAC 2559
Qy 2852 ACTCACGTCTCCAAATGAGCGCTTGGCGAGGAGAGAGCTGTGGCGGCTCATGGAGC 2911
Db 2560 ACAGACGTGTCCAAATGAGCGCTGTGTGCGCCGAGAGAGCTGTGGAGGCTCTGGAGC 2619
Qy 2912 GTGGCACTGATCCGACCTCTTCTGCTCGCTGAGGAGATTTTTCGCTGTGTGTA 2967
Db 2620 GTGGCTGTCTATCCGACACTCTTCCGCTCTGAGGACTACTTTGCTGTGTA 2675

RESULT 5
ID AA237096 standard; DNA; 4195 BP.
XX
AC AA237096;
XX
XX 27-MAR-2000 (first entry)
XX
DE DNA encoding de novo DNA cytosine methyltransferase Dnmt3b1.
XX
KW De novo DNA cytosine methyltransferase; Dnmt3b1; neoplastic disorder;
KW carcinoma; sarcoma; leukaemia; DNA methylation; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 269..2848
FT /tag=a
FT /product="de novo DNA cytosine methyltransferase
FT Dnmt3b1"
XX
PN W09967397-A1.
XX
XX 29-DEC-1999.
XX
XX 25-JUN-1999; 99NO-US14373.
XX
XX 25-JUN-1998; 98US-0090906.
XX
XX 24-JUN-1998; 98US-0093993.
XX
XX (GEO) GEN HOSPITAL CORP.
XX
XX Li E, Okano M, Xie S;
```

```
XX
DR WPI: 2000-106298/09.
DR P-PSDB; AA154056.
XX
XX New mouse and human polypeptides, useful to treat and diagnose
XX neoplastic disorders e.g. carcinomas, sarcomas and leukemias -
XX
XX Claim 9; Fig 1B; 114pp; English.
XX
XX The present sequence encodes a murine de novo DNA cytosine
XX methyltransferase designated Dnmt3b1. The Dnmt3b gene also produces,
XX through alternate splicing, at least two shorter isoforms of 840 and
XX 777 amino acid residues, termed Dnmt3b2 and Dnmt3b3 (sequences
XX not given in the specification). The polypeptides can be administered
XX therapeutically, especially by expressing encoding polynucleotides, to
XX treat diseases associated with DNA cytosine methyltransferase, such as
XX neoplastic disorders e.g. carcinomas, sarcomas and leukemias. They can
XX be used to diagnose or determine susceptibility to neoplastic disorders,
XX by assaying for polypeptide expression levels in mammalian cells/body
XX fluids. They are useful to screen for compounds inhibiting/activating the
XX polypeptide. The polypeptides can also be used for in vitro de novo
XX methylation of DNA. Such in vitro methylation may be used to direct or
XX regulate DNA expression in biological systems, e.g. recombinant DNA
XX methylated in vitro may be introduced into a cell/organism to increase
XX or decrease expression of a desired polypeptide for which the native DNA
XX is under-methylated or not methylated. The polypeptides can also be
XX used to produce antibodies which are useful to detect and purify the
XX polypeptide or therapeutically e.g. to treat neoplastic disorders. The
XX polynucleotides are useful to produce probes and primers which are
XX useful diagnostically.
XX
XX Sequence 4195 BP; 1061 A; 1059 C; 1123 G; 952 T; 0 other;
XX
XX Query Match 17.7%; Score 760.4; DB 21; Length 4195;
XX Best Local Similarity 63.7%; Pred. No. 6.8e-145;
XX Matches 1270; Conservative 0; Mismatches 661; Indels 63; Gaps 5;
Qy 1016 GAGCCGCTGGGCTCCGATGCTGGGGCAAGATGCCACCAAGAGGCGATGAGCGCCA 1075
Db 875 GATCAGGAGGTATGATATACACACAGGTGGATGCAGAGCATATATGGAGACAGACA 934
Qy 1076 GAGTACGAGGCGCGCGGCTTGGCATTTGGGAGCTGGTGGGGGAACTCGGGGGC 1135
Db 935 GAGTATCAGGATTAAGAGTTTGGATAGGTGACCTCTGTGGGGAAGATCAAGGGC 994
Qy 1136 TTCTCTGTGCGCCAGCGCCGATTTGTCTTGTGTGATGACGGGCGGAGCCGAGAGCT 1195
Db 995 TTCTCTGTGCGCTGCCCATGCTGCTGGAAGCCACCTCCAAGCAGAGGCCCATG 1054
Qy 1196 GAAGGCACCCGCTGGGTCACTGTTCCGAGACGCGCAAAATTTCTCAGTGGTGTGTTGAG 1255
Db 1055 CCGGAATCGCTGGGTACAGTGGTTTGTGTGATGTCGCAAGTTTCTGAGATCTCTGCTGAC 1114
Qy 1256 AAGCTGATCGCTGAGCTCGTTTTTGCAGTGGCTTCCACAGGCCACGTAACAACAGCAG 1315
Db 1115 AAATGTTGGTCTCTGGGCTGTTTCAGCCAGCAGCTTTAATCTGGCTACCTTCAATAAGCTG 1174
Qy 1316 CCATATCCGCAAGAGCCATCTACAGGTCTCTCAGGTGGCCAGCAGCGCGGGGAAG 1375
Db 1175 GTTCTTATAGGAAGCCATGTACACACTCTGGAGAAAGCCAGGGTTCGAGCTGGCAAG 1234
Qy 1376 CTGTTCCCGGTGTGCCACGACGAGGTGAGAGTGACACTGCCAAGCCCTGGAGGTGAG 1435
Db 1235 ACCTTCTCCAGCTC-----CTGGAGAGTCACTGGAGAGCAG 1273
Qy 1436 AACAAGCCCATGTTGAATGGGCGCTTCCAGCTTTCAGCCCTTTCGGCCCTTAAGGGCTG 1495
Db 1274 CTGAAGCCCATGCTGGAGTGGGCGCCACGGTGTCTCAAGCCCTACTGGGATCGAGGGCTC 1333
Qy 1496 GAGCC--ACCAGAAGAGAGAGAAATCCCTTACAAAGAGTGTACACGAGCATGTGGGTGG 1553
Db 1334 AAACCCCAACAAGAGCAACCCAGTGGTTAATAAGTCAAGGTGCTGCTGTTCAAGCAGTAGG 1393
```

OY	1554	AACCT-----GAGGAGAGCGCTACGACCACTCCAC	1587
DB	1334	AACCTTAGAACCCAGAGACCGGAGACAAAGTCGAGAGCGACACCAATGATCTGTCT	1453
OY	1588	AGCCAAAAGCCCGGAGAGACACAGCGGAGAGCCCAAGTCCAGAGATTA-----	1640
DB	1454	GCTTCTGATGTCCTCCCGCCACCGACGCGCTCAACAGCAAAATAGCTATGCGGGAAGGACGA	1513
OY	1641	-----TTATGAGCCACAGAGAGCGGCTGTGTACGAGGTGCGGAGATGCGCGAAC	1696
DB	1514	GCGGAGATGAGGAGAGCGCGAGAACGATGGCTTCGAAAGTCACCAACAAAGGGCAAT	1573
OY	1697	ATTGAGCATCTGCATCTCTCTGAGGAGGCTCAATGTTACCTGGAAACCCCTCTTC	1756
DB	1514	CTGGAAAGCGCTTTTGTCTCTGGAAAGAAAGACCTGTGTCTTCCACCCCTCTTT	1633
OY	1757	GTTGAGGAATAGTCCCAAACTCGAAGAACCTGTTCTGGAGTGTGCTGACAGTACGAC	1816
DB	1634	GAGGATGGGCTCTGTCAAGATGTCCGSGAATCGCTTCTAGAGCTCTTCAATGATAT	1693
OY	1817	GACGAGGCTACCAATCTCACTGACCACTGTCTGTGTGGGCGCGTGAAGGTCTCATGTG	1876
DB	1694	GAGGCGGCTATCAATCTCTATCGACCGTGTGTGAGGGCGGTGAATCTGTCTGTGTC	1753
OY	1877	GGAACCAACATCTGCTCAGGTGTTTGTGCTGAGATGTGTGACCTCTTGTGTGGGCGG	1936
DB	1754	AGTAACACAACTCTCTCAAGATGTTCTGTGTGAATGTCTGAGGTGTCTGTGTGGCGCA	1813
OY	1937	GGGGCTGCCAGGAGGCTATTAAGGAAGACCCCTGGAACTGTCACTGTGCGGAGCAAG	1996
DB	1814	GGCACAGCTGAGATGTCCCAAGCTGACAGAACCTTGAGCTGCTATATGTGCTCTCTAG	1873
OY	1997	GGTACTACGGGCTGTGCGGCGGAGAGACTGGCCCTCCGCTCCAGATGTTCTTC	2056
DB	1874	CGCTCCATGTGGGCTCTCCGACGAGAGAAAGATTGAAACATGGCGCTCGAAGACTTCTTC	1933
OY	2057	GCTATTAACCAAGC---AGGAATTTGACCCCTCCAAAGGTTTACCACCTGTCCAGCT	2113
DB	1934	ACTACTGATCTGTGACCTGGAAAGAAATTTGAGCACCCAAAGTTGATCCCAACATTTCTTCA	1993
OY	2114	GAGAAAGAGAGAGCCATCTCGGGTCTGTCTCTTTGATGGAATCGCTACAGGGCTCTCG	2173
DB	1994	GCCAAAGAGAGGCCCATTTAGAGTCTGTCTCTGTTGATGGAATTTGCAACGGGCTATCTTG	2053
OY	2174	GTGCTGAAGGACTTGGGCAATTCAGGTGACCGCTACATTTGCTCGAGAGGTGTGTAGGAC	2233
I	2054	GTGCTCAAGGAGTTGGGATTTAAAGTGGAAATATCACTTGTGCTCGAATCTGTGCAAG	2113
OY	2234	TCCATCACGCTGGGCAATGTGTGCGGACACAGAGGAAGATCATGTACGTGGGACGTCCG	2293
DB	2114	TCCATCTGTGTGGGAACGTTTAGCAATGAAGGCGCAATCAAAATATGTCAATGACCTCGG	2173
OY	2294	AGCGCACAGAGAGCAATTCAGAGATGGGGCCATTTGATCTGTGTATTTGGGGGAGT	2353
DB	2174	AAATACCAAGAAATATTTGAAAGATGGGCGCCGTGTGACTGTGTATTTGTGGAAGC	2233
OY	2354	CCCTGCAATGACCTCTCCATGTCAACCTGTCTGCAAGGGGCTCTACAGAGGCACTGGC	2413
DB	2234	CCATGCAATGATCTCTGTAAAGTAACTGTGCGCGAAAGTTTATATAGGGCGACAGAA	2293
OY	2414	CGGCTCTTCTTGAGCTTACCGGCTCTGCAATGATCGCGGCGCAAGAGGAGATAT	2473
DB	2294	AGGCTCTTCTTGAGTTTACCTGTGTAATTAATACCGGCGCCCAAGAGGGCGCAAC	2353
OY	2474	CGCCCTCTCTTGAAGTGTGGAATGTGTGAGCCATGGAAGTGAATACCAAGAAAGAC	2533
DB	2354	CGTCCATTTCTTGAGATGTTCGAATGTGTGACCAATGAAGTGAATACCAAGAAAGAC	2413
OY	2534	ATTCGCGAATTTCTGAGTCCAAACCTGTATGATTTGATGCCAAAGAAATGTCACTGCA	2593
DB	2414	ATTCMAAATTCCTGGAGATGTAAACCGAGTATATGATGATGCCATCAAGGTTGTCTGTCT	2473
OY	2594	CAGAGGCGCCGCTACTTCTGGGGTAACTTTCCGGATGTAAACAGCGCGTTTGGCATCACT	2653

CC of cells from a patient. (I) is useful for determining the presence of  
CC colon cancer in a cell or tissue type, for determining the presence or  
CC state of other type of cancer, in antisense therapy, to generate  
CC macroarrays on a solid surface, to identify a chromosome on which the  
CC corresponding gene resides, and in tissue profiling, forensic, genetic  
CC analysis, mapping and diagnostic applications. (I) can be used to raise  
CC antibodies, and to screen for peptide analogues and antagonists.  
XX  
SQ Sequence 709 BP; 163 A; 177 C; 193 G; 168 T; 8 other;

Query Match 14.3%; Score 613; DB 24; Length 709;  
Best Local Similarity 94.5%; Pred. No. 3.6e-115;  
Matches 664; Conservative 0; Mismatches 33; Indels 6; Gaps 3;

QY 2275 GTAGCTCGGGAGCTCCGAGGTGACACAGAGCATATCCAGAGTGGGGCCCATTCGA 2334  
DB 1 GGAGCTCGGGAGCTCCGAGGTGACACAGAAATATCCAGAGTGGGGCCCATTCGA 60  
QY 2335 TCTGGTATTTGGGGGAGTCCCTGCAATGACCTCTTCATCGTCAACCCCTGCTCGCAAGG 2394  
DB 61 TCTGGTATTTGGGGGAGTCCCTGCAATGACCTCTTCATCGTCAACCCCTGCTCGCAAGG 120  
QY 2395 CCTTACGAGGCACTGGCGGCTCTTCTTTGAGTTCTACCGCTCTGCAATGATGCGG 2454  
DB 121 CCTTACGAGGCACTGGCGGCTCTTCTTTGAGTTCTACCGCTCTGCAATGATGCGG 180  
QY 2455 GCCCAAGGAGAGATGATGCGCCCTCTTCTGCTCTTTGAGAATGTGGTGGCCATGGG 2514  
DB 181 GCCCAAGGAGAGATGATGCGCCCTCTTCTGCTCTTTGAGAATGTGGTGGCCATGGG 240  
QY 2515 GCTTAGTGAACAGAGGACATCTCCGATTTCTCGAGTTCACACCCCTGTCATGATGTC 2574  
DB 241 GCTTAGTGAACAGAGGACATCTCCGATTTCTCGAGTTCACACCCCTGTCATGATGTC 300  
QY 2575 CAAGAGTGTGAGTGCACACAGGCGCGTACTTCTGGGTAACTTCCCGGTATGAA 2634  
DB 301 CAAGAGTGTGAGTGCACACAGGCGCGTACTTCTGGGTAACTTCCCGGTATGAA 360  
QY 2635 CAGCGGTGGCATCCACTGTGATGATAGCTGAGCTGCAGAGTGTGCGACATGG 2694  
DB 361 CAGCGGTGGCATCCACTGTGATGATAGCTGAGCTGCAGAGTGTGCGACATGG 420  
QY 2695 CAGATAGCCAAAGTTCAGAAAGTGGAGCAATTTACTAGAGTCAAACTCCATAAGCA 2754  
DB 421 CAGATAGCCAAAGTTCAGAAAGTGGAGCAATTTACTAGAGTCAAACTCCATAAGCA 480  
QY 2755 GGGCAAGACCAAGCATTTTCTGTCTTCATGAATGAGAAAGAGGACATCTTATGT-GCA 2813  
DB 481 GGGCAAGACCAAGCATTTTCTGTCTTCATGAATGAGAAAGAGGACATCTTATGTGGCA 540  
QY 2814 CTGAAATGGAAGAGGTATTTGGTTTCCAGTCCACTATAGTACGTCTCCAACTAGGCC 2873  
DB 541 CTGAAATGGAAGAGGTATTTGGTTTCCAGTCCACTATAGTACGTCTCTCAACATGACC 600  
QY 2874 GCTTGGCGAGGAGACTGCTGGGCGGTCA-TGGAGCGTGCA---GTCAATCCGCA 2928  
DB 601 GCTTGGCGAGGAGAACTGCTGGGCGGTCAATTTGGACCGTGCCAAATTCATCCNCACT 660  
QY 2929 CCTCTTCGCTCCGCTGAAGGATATTTTGGTGTGTGTTAGGG 2971  
DB 661 TTTTNGTTCGCTGAAGGATATTTTGGTGTGTGTTAGGG 703

## RESULT 7

AAH33860  
ID AAH33860 standard; cDNA; 1060 BP.

XX  
AC  
XX  
XX  
AAH33860;

DT 03-SEP-2001 (first entry)

XX Human colon cancer antigen encoding cDNA SEQ ID NO:916.

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200122920-A2.  
XX  
XX 05-APR-2001.  
XX  
XX 28-SEP-2000; 2000WO-US26524.  
XX  
XX 29-SEP-1999; 99US-0157137.  
XX  
XX 03-NOV-1999; 99US-0163280.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX  
XX WPI; 2001-235357/24.  
XX  
XX P-PSDB; AAG74429.  
XX  
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
XX useful for preventing, diagnosing and/or treating colorectal cancers -  
XX  
XX Claim 1; Page 2829; 9803pp; English.

XX  
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
XX cancer-associated nucleic acid molecules (N) and proteins (P), where  
XX the proteins are collectively known as colon cancer antigens. The colon  
XX cancer antigens have cytostatic activity and can be used in gene  
XX therapy and vaccine production. N and P may be used in the prevention,  
XX diagnosis and treatment of diseases associated with inappropiate P  
XX expression. For example, N and P may be used to treat disorders  
XX associated with decreased expression by rectifying mutations or deletions  
XX in a patient's genome that affect the activity of P by expressing P.  
XX Inactive proteins or to supplement the patient's own production of P.  
XX Additionally, N may be used to produce the colon cancer-associated P,  
XX by inserting the nucleic acids into a host cell and culturing the cell  
XX to express the proteins. N and P can be used in the prevention, diagnosis  
XX and treatment of colorectal carcinomas and cancers. AAH37195 to AAH37204  
XX and AAH7789 represent sequences used in the exemplification of the  
XX present invention.

XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
XX missing at time of publication, meaning no sequences are present for  
XX SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 1060 BP; 267 A; 262 C; 297 G; 227 T; 7 other;

Query Match 12.8%; Score 548.2; DB 22; Length 1060;  
Best Local Similarity 72.6%; Pred. No. 6.7e-102;  
Matches 700; Conservative 4; Mismatches 260; Indels 0; Gaps 0;

QY 2004 ACGGGCTGTGGGCGGAGAGACTGGCCCTCCCGCTCCAGATGTTCTTCGCTAATA 2063  
DB 18 ATGGCGTCTCTGGCGCGGAGAGACTGGAACTGGACGTGCGCTGCAGGCTTCTTCAACAGT 77  
QY 2064 ACCACGACAGGAATTTGACCCCTCCAAAGTTTACCCACCTGTCCAGCTGAGAGAGGA 2123  
DB 78 ACACGGGCTTGAATACGAAGCCGCCAGCTGACCTGCCATTCGCCGACCCGAGGC 137  
QY 2124 AGCCCATCGGGTGTCTCTCTCTTTGATGATGATGCTACAGGGCTCTCTGGTCTGAGG 2183  
DB 138 GGCCCATTCGAGTCTCTGTCTTTGATGATGATGCTACAGGGCTCTCTAGTCTCAAG 197  
QY 2184 ACTTGGCATTTCAAGTGGACCGCTACATTTGCCCTCGAGGTGTGTGAGACTCCATCAGG 2243  
DB 198 AGTTGGCATTAAGGTAGGAAGTACGTCTCTTCTGAAAGTGTGTGAGAGTCCATTGCTG 257  
QY 2244 TGGGCATGTGGGCAACAGGGGAAGATCATGTACGTGGGGAGCTCCGACGCTCACAC 2303  
DB 258 TTGGAACCGTGAAGCAGGAGGGAATATCAAAATACGTGAACGATGTGAGGAACATCAAA 317  
QY 2304 AGAAGCATATCCAGGAGTGGGGCCCATTCGATCTGTTGATTTGGGGGAGTCCCTGCAATG 2363



PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251472.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2000US-0254097.  
PR 05-JAN-2001; 2000US-0255678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
PI WPI; 2001-465570/50.  
XX P-PSDB; AAM94301.  
XX Isolated nucleic acid molecule encoding a reproductive system antigen  
PT is used in preventing, treating or ameliorating a medical condition -  
XX P-PSDB; AAM94301.  
PS Claim 1; SEQ ID NO 272; 1297pp + Sequence Listing; English.  
XX The present invention provides the protein and coding sequences of a

CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a coding sequence of the  
CC invention.  
XX  
SQ Sequence 622 BP; 154 A; 161 C; 189 G; 112 T; 6 other;  
Query Match 11.8%; Score 504.6; DB 22; Length 622;  
Best Local Similarity 98.0%; Pred No. 4.1e-93;  
Matches 540; Conservative 2; Mismatches 6; Indels 3; Gaps 3;  
Qy 1471 CACGCTTCTGGCCCTAAGGCTGAGGCACACAGAAAGAGAGATCCCTACAAAGA 1530  
Db 2 CACGCTTCTGGCCCTAAGGCTGAGGCACACAGAAAGAGAGATCCCTACAAAGA 61  
Qy 1531 AGTGACACGGACATGTGGGTGGAACCTGAGCAGCTGCTACGACACCTCCACGAGC 1590  
Db 62 AGTGACACGGACATGTGGGTGGAACCTGAGCAGCTGCTACGACACCTCCACGAGC 121  
Qy 1591 CAAAAGCCCGGGAAGAGACAGCGGAGAGCCCAAGGTCAAGGAGATTATTGATGAGCG 1650  
Db 122 CAAAAGCCCGGGAAGAGACAGCGGAGAGCCCAAGGTCAAGGAGATTATTGATGAGCG 181  
Qy 1651 CACAAGAGAGCGGCTGGTGTACGAGGTGCGGAGAGAGTCCCGGAACATTGAGGACATCTG 1710  
Db 182 CACAAGAGAGCGGCTGGTGTACGAGGTGCGGAGAGAGTCCCGGAACATTGAGGACATCTG 241  
Qy 1711 CATCTCTGTGGAGCCTCAATGTTACCTGGAAACACCCCTCTTCGTTGGAGGAATGTG 1770  
Db 242 CATCTCTGTGGAGCCTCAATGTTACCTGGAAACACCCCTCTTCGTTGGAGGAATGTG 301  
Qy 1771 CCAAACTGCAAGAACTGCTTCTGGAGTGTGGTACGATGACGAGACGCGCTACCA 1830  
Db 302 CCAAACTGCAAGAACTGCTTCTGGAGTGTGGTACGATGACGAGACGCGCTACCA 361  
Qy 1831 GTCCTACTGCACCATCTGCTGTGGGGGCGGTGAGGTGCTCATGTGCGGAAACAACTG 1890  
Db 362 GTCCTACTGCACCATCTGCTGTGGGGGCGGTGAGGTGCTCATGTGCGGAAACAACTG 421  
Qy 1891 CTGCAAGTGCTTTTTCGTTGAGTGTGGAACCTCTTGTGGGGCGGGGGCTGCCAGGC 1950  
Db 422 CTGCAAGTGCTTTTTCGTTGAGTGTGGAACCTCTTGTGGGGCGGGGGCTG-GCAAGC 480  
Qy 1951 AGCCATTAAAGGAAGACCCCTGGAACCTGCTACATGTGCGGGCACAAGGCTACCTACGGCT 2010  
Db 481 AGCCATTAAAGGAAGACCCCTGGAACCTGCTACATGTGCGGGCACAAGGCTACCTACGGCT 538  
Qy 2011 GCTGCGGCGGC 2021  
Db 539 GCTGCGGCGGC 549  
RESULT 9  
AAHL4008  
ID AAHL4009 standard; cDNA; 3017 BP.  
XX  
AC AAHL4009;  
XX  
XX 26-JUN-2001 (first entry)  
DT Human cDNA sequence SEQ ID NO:11099.  
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX  
XX 07-FEB-2001.  
PD  
XX 28-JUL-2000; 2000EP-0116126.  
XX  
XX 29-JUL-1999; 99JP-0248036.  
PR





Db 1440 TGTGCACTACA 1450

RESULT 10  
AAS86165  
ID AAS86165 standard; cDNA; 711 BP.  
XX  
AC AAS86165;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #21969.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
FN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG21978.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 1; SEQ ID No 21969; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX disorders involving aberrant protein expression or biological activity.  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. AAS64197-AAS94564 represent novel human  
XX diagnostic coding sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 711 BP; 181 A; 164 C; 199 G; 167 T; 0 other;

Query Match 8.0%; Score 144.2; DB 23; Length 711;  
Best Local Similarity 75.0%; Pred. No. 2e-60;  
Matches 430; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

Qy 2135 GTCTCTCTCTTTTGTGATCGTACAGGGCTCTGTGCTGAAGACTTGGCAATT 2194  
Db 1 GTCTCTCTCTTTTGTGATCGTACAGGGCTCTGTGCTCTCAAGAGTGGGCATA 60

Qy 2195 CAGGTGGACCGCTACATTGCTCGAGGTGTGTGAGGACTCCATCAGGGTGGCATGGTG 2254  
Db 61 AAGGTAGGAAAAGTACGTCGCTTCTGAAGTGTGTGAGGAGTCCATTGCTGTGGAACCGTG 120  
Qy 2255 CGGCACACAGGGGAAGATCATGTACGTCCGGGAGCGTCCCGACGCTCACACAGAAAGCATATC 2314  
Db 121 AAGCACAGAGGGGAATATCAAAATACGTGAACGACGTGAGGAACATCAAAAAGAAAATATT 180  
Qy 2315 CAGGAGTGGGCGCCATTCGATCTGGTATTGGGGGACGTCCCTCAATGACCTCTCCATC 2374  
Db 181 GAAGAATGGGCGCCATTTGACTTGGTATTGGCGGAAGCCCATCAACGATCTCTCAAT 240  
Qy 2375 GTCAACCTGCTCTCGAAGGCGCTCTACAGAGGACCTGCGCGCTCTCTTTGAGTTCTAC 2434  
Db 241 GTGATTCAGCAGAGAAAGCCCTGTATGAGGGTACAGGCCGCTCTCTTCGAATTTTAC 300  
Qy 2435 CGCTCTCTGATGATGCGCGGCCCAAGAGGAGGAGATGATCCGCCCTCTCTCTGGCTCTTT 2494  
Db 301 CACCTGCTGAATTTACTCACGCCCAAGAGGGTGTATGACCGCGCTCTCTCTGGATGTTT 360  
Qy 2495 GAGAATGTGGTGGCCATGGCGTTACTGACAAGAGGGACATCTCGCATTTCTCGAGTCC 2554  
Db 361 GAGAATGTGTAGCCATGAAGGTGTGGCGACAAGAGGGACATCTCACGGTTCTCGAGTGT 420  
Qy 2555 AACCTCTGTGATGATGATGCAAGAAAGTGTGAGTGCACACAGGGCCCGCTACTTTCTGG 2614  
Db 421 AATCCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
Qy 2615 GGTAACTTCCCGGTATGAACAGCGCGTTGGCATCCACTGTGAATGATAAGCTGGAGCTG 2674  
Db 481 GGCACCTTACCCGGGATGAACAGCGCCGTTGATGATGATGATGATGATGATGATGATGATG 540  
Qy 2675 CAGGAGTCTCTGGAGCATGGCAGGATAGCCAG 2707  
Db 541 CAGGACTCTTGGAAATACAATAGATAGCCNAG 573

RESULT 11  
AAF80537  
ID AAF80537 standard; cDNA; 2077 BP.  
XX  
AC AAF80537;  
XX  
DT 08-JUN-2001 (first entry)  
XX  
DE Receptor #25 partial coding sequence.  
XX  
KW Probe; microarray; cancer; immunopathology; neuropathology; ss.  
XX  
OS Balaena mysticetus.  
XX  
PN US6183968-B1.  
XX  
PD 06-FEB-2001.  
XX  
PF 25-MAR-1999; 99US-0276531.  
XX  
PR 27-MAR-1998; 98US-0079677.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Bandman O, Lal P, Hillman JL, Yue H, Reddy R, Guegler KJ;  
PI Baughn MR;  
XX  
XX WPI; 2001-201999/20.  
XX  
XX Composition having probes which comprise part of gene sequence encoding  
XX proteins associated with cell proliferation useful as hybridizable  
XX array elements in Microarrays to monitor expression of target  
XX polynucleotide -  
XX  
XX Claim 1; Columns 93-96; 104pp; English.

CC The present invention relates to a composition comprising several  
CC polynucleotide probes. Probes can be derived from the present sequence.  
CC The probes are immobilised and are preferably useful as hybridisable  
CC array elements in a microarray for monitoring the expression of several  
CC polynucleotides. The microarray can be used in the diagnosis of cancers  
CC such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma  
CC and tetracarcinoma etc., immunopathology such as AIDS, Addison's  
CC disease, adult respiratory distress syndrome, allergies, anaemia, asthma,  
CC atherosclerosis and bronchitis etc., neuropathology such as Alzheimer's  
CC disease, amnesia, amyotrophic lateral sclerosis, bipolar disorder,  
CC cataracta and cerebral neoplasms etc. The microarray can also be used to  
CC investigate an individual's predisposition to a disease such as cancer,  
CC immunopathology or neuropathology. Also, the microarray can be used for  
CC investigating cellular response to infection, drug treatment etc. The  
CC microarray can be used for diagnostics, prognostics and treatment  
CC regimens, drug discovery and development, toxicological and  
CC carcinogenicity studies, forensics, pharmacogenomics etc. The array can  
CC also be used for monitoring disease progression.

SY Sequence 2077 BP; 539 A; 485 C; 477 G; 567 T; 9 other;

Query Match 7.6%; Score 324.6; DB 22; Length 2077;

Best Local Similarity 74.8%; Pred. No. 3e-56; Mismatches 139; Indels 1; Gaps 1;

Matches 416; Conservative 0; Mismatches 139; Indels 1; Gaps 1;

QY 2412 GCCGGCTCTTCTTGAAGTTCACCGCCCTTCATGATGCGCGCCCAAGAGGAGATG 2471  
DB 62 GCGCGCTCTTCTTGAAGTTCACCGCCCTTCATGATGCGCGCCCAAGAGGAGATG 120

QY 2472 ATGCGCCCTTCTTGAAGTTCACCGCCCTTCATGATGCGCGCCCAAGAGGAGATG 2531  
DB 121 ACCGCGCTCTTCTTGAAGTTCACCGCCCTTCATGATGCGCGCCCAAGAGGAGATG 180

QY 2532 ACATCTCGCGATTCCTGAGTCCAAACCTGATGATGATGATGATGATGATGATGATGATG 2591  
DB 181 ACATCTCGCGATTCCTGAGTCCAAACCTGATGATGATGATGATGATGATGATGATGATG 240

QY 2592 CACACAGGCGCGCTTCTGAGTCCAAACCTGATGATGATGATGATGATGATGATGATGATG 2651  
DB 241 CTCACAGGCGCGCTTCTGAGTCCAAACCTGATGATGATGATGATGATGATGATGATGATG 300

QY 2652 CTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2711  
DB 301 CAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360

QY 2712 GCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2771  
DB 361 AGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420

QY 2772 TTCTGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2831  
DB 421 TTCTGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480

QY 2832 TTGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2891  
DB 481 TTGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540

QY 2892 TGTCTGCGCGCTTCTGAGTCCAAACCTGATGATGATGATGATGATGATGATGATGATG 2951  
DB 541 TGTCTGCGCGCTTCTGAGTCCAAACCTGATGATGATGATGATGATGATGATGATGATG 600

QY 2952 ATTTGGCTGTGTGA 2967  
DB 601 ACTTTGATGTGATTA 616

RESULT 12

AAT21884

ID AAT21884 standard; cDNA to mRNA; 301 BP.

XX AAT21884;

AC AAT21884;

XX 14-AUG-1996 (first entry)

XX Human gene signature HUMGS03426.

XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;

XX human; cloning; mapping; non-biased library; diagnosis; detection;

XX cell typing; abnormal cell function; ss.

XX

OS Homo sapiens.

XX

PN W09514772-A1.

XX

PD 01-JUN-1995.

XX

PF 11-NOV-1994; 94MO-JF01916.

XX

PR 12-NOV-1993; 93JP-0355504.

XX

PA (MATS/) MATSUBARA K.

XX (OKUB/) OKUBO K.

PI Matsubara K, Okubo K;

XX

DR WPI; 1995-206931/27.

XX

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.

PT for diagnosis of abnormal cell function, by preparing cDNA that

PT reflects relative abundance of corresp. mRNA in specific human

PT tissues

PS Claim 1; Page 991-992; 2245BP; Japanese.

XX

CC A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in AAT19001-T26837 and which is able to hybridise to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost

CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library

CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be

CC determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.

XX

SQ Sequence 301 BP; 96 A; 65 C; 63 G; 66 T; 11 other;

Query Match 6.6%; Score 285.2; DB 16; Length 301;

Best Local Similarity 95.3%; Pred. No. 1.5e-48; Mismatches 14; Indels 0; Gaps 0;

Matches 287; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 3519 GATCAGATAGAGACACAGAGGAGCGAAGAGAGAGACACTCAGCGGACGATTC 3578  
DB 1 GATCAGATAGAGACACAGAGGAGCGAAGAGAGAGACACTCAGCGGACGATTC 60

QY 3579 CTCCAGCCTAGCTGTGTGCGACAGCCTTCTGTCAGCAAAAGAGACCACT 3638  
DB 61 CTCCAGCCTAGCTGTGTGCGACAGCCTTCTGTCAGCAAAAGAGACCACT 120

QY 3639 TAGCAGAGGAG 3698  
DB 121 TAGCAGAGGAG 180

QY 3699 CCACACAGAGAACTTGAAGAAATCAGTTCTAGAACCGCTGTACCTCTTGTTTACA 3758  
DB 181 CCACACAGAGAACTTGAAGAAATCAGTTCTAGAACCGCTGTACCTCTTGTTTACA 240

QY 3759 GTTTAT 3818  
DB 241 GTTTAT 300

```
Oy 3819 A 3819'
Db 301 A 301

RESULT 13
ABN97389/c
ID ABN97389 standard; DNA; 273 BP.
XX AC ABN97389;
XX DT 13-AUG-2002 (first entry)
XX DE Gene #3887 used to diagnose liver cancer.
XX KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
XX KW metastatic liver tumour; cytostatic; expression profile; disease state;
XX KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX OS Homo sapiens.
XX PN WO200229103-A2.
XX PD 11-APR-2002.
XX PF 02-OCT-2001; 2001WO-US30589.
XX PR 02-OCT-2000; 2000US-237054P.
XX PA (GENE-) GENE LOGIC INC.
XX PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI; 2002-426119/45.
XX DI Diagnosing and detecting the progression of liver cancer.
XX PT hepatocellular carcinoma or metastatic liver tumor in a patient, in a
XX PT liver tissue sample.
XX PS Claim 1; SEQ ID NO 3887; 298pp; English.
XX SS The invention relates to a novel method for diagnosing and detecting the
XX CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX CC tumour in a patient, and differentiating metastatic liver cancer from
XX CC hepatocellular carcinoma in a patient, involving detecting the level of
XX CC expression of two or more genes represented in ABN97455 in a
XX CC tissue sample. The method of the invention has hepatotropic, and
XX CC cytostatic activity. The method is useful for diagnosing and detecting
XX CC the progression of liver cancer, hepatocellular carcinoma and metastatic
XX CC liver carcinoma in a patient. The method is useful for identifying
XX CC expression profiles which serve as useful diagnostic markers as well as
XX CC markers that can be used to monitor disease states, disease progression,
XX CC drug toxicity, drug efficacy and drug metabolism.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pat_sequences.
XX SQ Sequence 273 BP; 59 A; 74 C; 62 G; 78 T; 0 other;

Query Match 6.4%; Score 273; DB 24; Length 273;
Best Local Similarity 100.0%; Pred. No. 4.5e-46;
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4002 CAGGTTCTCCGCAAAAGGGGCTAGATGAGGTGGTGGGCCCGCGGACAGCTGAGA 4061
Db 273 CAGGTTCTCCGCAAAAGGGGCTAGATGAGGTGGTGGGCCCGCGGACAGCTGAGA 214

Oy 4062 GTGGGATTTCACGACTCATCAATACCCCTTTGATTGTTTCTAAAGGAGACTCCCT 4121
Db 213 GTGGGATTTCACGACTCATCAATACCCCTTTGATTGTTTCTAAAGGAGACTCCCT 154

Oy 4122 CGGCAAGATGGCAGAGGTAACGGAGTCTTCAGGCCCAAGTTTCTCACTTTAGCCAATTGCA 4181
```

```
Db 153 CGGCAAGATGGCAGAGGTAACGGAGTCTTCAGGCCCAAGTTTCTCACTTTAGCCAATTGCA 94
Oy 4182 GGGCTCCTTGTGGTGGGATCAGAACTAATCCAGAGTGTGGGAAAGTGACAGTCAAAACCC 4241
Db 93 GGGCTCCTTGTGGTGGGATCAGAACTAATCCAGAGTGTGGGAAAGTGACAGTCAAAACCC 34
Oy 4242 CACCTGGAGCAAAATAAAAAAACATACAAAACGT 4274
Db 33 CACCTGGAGCAAAATAAAAAAACATACAAAACGT 1

RESULT 14
ABL64228/c
ID ABL64228 standard; DNA; 273 BP.
XX AC ABL64228;
XX DT 15-MAY-2002 (first entry)
XX DE Stomach cancer related gene sequence SEQ ID NO:2565.
XX KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
XX KW cytostatic; gene therapy; antineoplastic; wilm's tumour; adenocarcinoma;
XX KW gene; ds.
XX OS Homo sapiens.
XX PN WO200194629-A2.
XX PD 13-DEC-2001.
XX PF 30-MAY-2001; 2001WO-US10838.
XX PR 05-JUN-2000; 2000US-209473P.
XX PR 05-JUN-2000; 2000US-209511P.
XX PR 18-SEP-2000; 2000US-233113P.
XX PR 18-SEP-2000; 2000US-233617P.
XX PR 20-SEP-2000; 2000US-234009P.
XX PR 20-SEP-2000; 2000US-234034P.
XX PR 20-SEP-2000; 2000US-234052P.
XX PR 22-SEP-2000; 2000US-234509P.
XX PR 22-SEP-2000; 2000US-234567P.
XX PR 25-SEP-2000; 2000US-234923P.
XX PR 25-SEP-2000; 2000US-234924P.
XX PR 25-SEP-2000; 2000US-235077P.
XX PR 25-SEP-2000; 2000US-235082P.
XX PR 25-SEP-2000; 2000US-235134P.
XX PR 25-SEP-2000; 2000US-235280P.
XX PR 26-SEP-2000; 2000US-235637P.
XX PR 26-SEP-2000; 2000US-235711P.
XX PR 27-SEP-2000; 2000US-235720P.
XX PR 27-SEP-2000; 2000US-235840P.
XX PR 27-SEP-2000; 2000US-235843P.
XX PR 28-SEP-2000; 2000US-236038P.
XX PR 28-SEP-2000; 2000US-236032P.
XX PR 28-SEP-2000; 2000US-236033P.
XX PR 28-SEP-2000; 2000US-236034P.
XX PR 28-SEP-2000; 2000US-236109P.
XX PR 28-SEP-2000; 2000US-236111P.
XX PR 29-SEP-2000; 2000US-236842P.
XX PR 29-SEP-2000; 2000US-236891P.
XX PR 02-OCT-2000; 2000US-237172P.
XX PR 02-OCT-2000; 2000US-237173P.
XX PR 02-OCT-2000; 2000US-237278P.
XX PR 02-OCT-2000; 2000US-237294P.
XX PR 02-OCT-2000; 2000US-237295P.
XX PR 02-OCT-2000; 2000US-237316P.
XX PR 03-OCT-2000; 2000US-237425P.
XX PR 03-OCT-2000; 2000US-237598P.
XX PR 03-OCT-2000; 2000US-237604P.
```

PR 03-OCT-2000; 2000US-237606P.  
 PR 03-OCT-2000; 2000US-237608P.  
 PR 01-NOV-2000; 2000US-244867P.  
 PR 01-NOV-2000; 2000US-245084P.  
 XX  
 PA (AVALON) AVALON PHARM.  
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S,  
 PI Soppet DR, Weaver Z;  
 DR WPI; 2002-188264/24.  
 XX  
 PT Screening for anti-neoplastic agent involves exposing cells to a  
 PT chemical agent to be tested for anti-neoplastic activity, and  
 PT determining a change in expression of a gene of a signature gene set  
 XX  
 PS Claim 1; SEQ ID 2565; 44pp; English.  
 CC The present invention describes a method (M1) for screening for an  
 CC anti-neoplastic agent. The method involves exposing cells to a chemical  
 CC agent to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (I) of a signature gene set, where (I)  
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
 CC to ABL70110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening  
 CC an anti-neoplastic agent, and can be used for producing a product which  
 CC is the data collected with respect to the anti-neoplastic agent as a  
 CC result of M1, and the data is sufficient to convey the chemical  
 CC structure and/or properties of the agent. M1 can be used in the  
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
 CC carcinoma, papillary carcinoma and Wilms' tumour.  
 XX  
 SQ Sequence 273 BP; 59 A; 74 C; 62 G; 78 T; 0 other;  
 Query Match 6.4%; Score 273; DB 24; Length 273;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-46;  
 Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4002 CAGGTTCTCCCTCCCAAAAAGGGGGCTAGATGAGTGGTGGGCCCTGACAGCTGAGA 4061  
 DB 273 CAGGTTCTCCCTCCCAAAAAGGGGGCTAGATGAGTGGTGGGCCCTGACAGCTGAGA 214  
 CC 4062 GTGGGATTCATCCAGACTCATGCAATACCTTTGATTGTTTCTAAAGAGACCTCCCT 4121  
 DB 213 GTGGGATTCATCCAGACTCATGCAATACCTTTGATTGTTTCTAAAGAGACCTCCCT 154  
 QY 4122 CGGCAAGATGGCAGAGGGTACGAGAGTCTCAGGCCAGTTTCTCACTTAGCCAAATTGGA 4181  
 DB 153 CGGCAAGATGGCAGAGGGTACGAGAGTCTCAGGCCAGTTTCTCACTTAGCCAAATTGGA 94  
 QY 4182 GGGCTCTCTTGTGGTGGATGAGACTAATCAGAGTGTGGGAAAGTACAGCTCAAAACC 4241  
 DB 93 GGGCTCTCTTGTGGTGGATGAGACTAATCAGAGTGTGGGAAAGTACAGCTCAAAACC 34  
 QY 4242 CACCTGAGCAATTAATAAACAATACAAACGT 4274  
 DB 33 CACCTGAGCAATTAATAAACAATACAAACGT 1  
 RESULT 15  
 ID ABA20255 standard; DNA; 546 BP.  
 XX ABA20255;  
 AC  
 XX  
 DT 23-JUN-2002 (first entry)  
 XX  
 DE Human nervous system related polynucleotide SEQ ID NO 12586.  
 XX

KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; antineoplastic; anti-HIV; antibacterial; vaccine;  
 KW antiparkinsonian; antischizophrenic; antianemic; antidiabetic; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
 KW antiallergic; antidiabetic; anticonvulsant; antifungal;  
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotoxic; gene therapy; vaccine; ds.  
 XX Homo sapiens.  
 XX  
 PN WO200159063-A2.  
 PD 16-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01334.  
 XX  
 PR 31-JUN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226688.  
 PR 22-AUG-2000; 2000US-0227069.  
 PR 23-AUG-2000; 2000US-0227182.  
 PR 30-AUG-2000; 2000US-0228824.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229609.  
 PR 05-SEP-2000; 2000US-0229613.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231245.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0234984.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 20-OCT-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246533.  
PR 08-NOV-2000; 2000US-0246509.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250191.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.

PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-541565/60.  
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating nervous system  
XX cancers and metastases -  
XX Disclosure; SEQ ID NO 12586; 1701pp + Sequence Listing; English.  
XX The invention relates to novel genes (ABAI1004-ABA21534) and proteins  
XX (ABBI4678-ABBI8001) useful for preventing, treating or ameliorating  
XX medical conditions e.g. by protein or gene therapy. The genes are  
XX isolated from a range of human tissues disclosed in the specification.  
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful  
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone  
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
XX and parasitic infections.  
XX Note: The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 546 BP; 133 A; 137 C; 171 G; 105 T; 0 other;

Query Match 5.8%; Score 254.6; DB 22; Length 546;  
Best Local Similarity 98.2%; Pred. No. 3.3e-42;  
Matches 268; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 406 GGTGGAAGCGGTGACGCGCAAGGACCTCGGTGATCTCCAAGTCCCATCCATGCC 465  
Db 2 GGTGGAAGCGGTGACGCGCAAGGACCTCGGTGATCTCCAAGTCCCATCCATGCC 61  
Qy 466 CCAGGACTCAGGCGCTCAGAGCTATTACCAATGGGACTTTGGAGAAGCGGAGTGACC 525  
Db 62 CCAGGACTCAGGCGCTCAGAGCTATTACCAATGGGACTTTGGAGAAGCGGAGTGACC 121  
Qy 526 CCAGCCAGAGGAGGCGGAGCCCTGCTGGGGGCGAGAAGGCGGGCCCGAGCAGAGGAGGA 585  
Db 122 CCAGCCAGAGGAGGCGGAG-CCTGCTGGGGGCGAGAAGGCGGGCTCCCGAGCAGAGGAGGA 180  
Qy 586 GGGTGCAGCTGAGAGCCCTGCTGAAGCCTCAAGAGCAGTGGGAAATGGCTGCTGCACCCC 645  
Db 181 GGGTGCAGCTGAGAGCCCTGCTGAAGCCTCAAGAGCAGTGGGAAATGGCTGCTGCACCCC 240  
Qy 646 CAAAGGAGGCGCGAGGAGCCCTGCTGAGAGCGGG 678  
Db 241 CAAAGGAGGCGCGAGGAGCCCTGCTGAGAGCGGG 273

Search completed: November 22, 2002, 09:12:08  
Job time : 616.737 secs